R E P O R T   R E S U M E S

ED 010 609   24

THE PROBLEM OF CLASSIFYING MEMBERS OF A SINGLE POPULATION INTO GROUPS. STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA, PART II.

BY- SAW, J.G.   FLORA, R.E.
VIRGINIA POLYTECHNIC INST., BLACKSBURG

REPORT NUMBER CRP-1132-PT-2   PUB DATE   65
REPORT NUMBER BR-5-0224
EDRS PRICE MF-$0.18 HC-$3.36   84P.

DESCRIPTORS- *MATHEMATICAL MODELS, *EDUCATIONAL RESEARCH, MATHEMATICAL APPLICATIONS, CALCULATION, MEASUREMENT TECHNIQUES, *PREDICTIVE ABILITY (TESTING), *PREDICTIVE MEASUREMENT, PREDICTIVE VALIDITY, *STATISTICAL ANALYSIS, BLACKSBURG, VIRGINIA

AN ALTERNATIVE MATHEMATICAL APPROACH TO THE PROBLEM OF CLASSIFICATION OF AN INDIVIDUAL BY VECTOR SCORES WAS PROPOSED FOR THOSE CASES WHEN THE USUAL CLASSICAL MODEL DOES NOT HOLD. THE NEW APPROACH ASSUMED THAT, WITHIN A GROUP HOMOGENEOUS IN REGARD TO BACKGROUND, TWO DISTINCT POPULATIONS OF PASS AND FAIL DO NOT EXIST, BUT RATHER A CONTINUOUS SPECTRUM. MAKING USE OF THIS NEW MODEL, THE INVESTIGATORS WERE ABLE TO PROPOSE A METHOD OF EVALUATING THE VECTOR SCORES ON MATHEMATICS DEPARTMENT ADMISSIONS TESTS IN RESPECT TO THEIR VALUE AS INDICATORS OF THE ULTIMATE WORTH OR PERFORMANCE OF THE INDIVIDUAL. THE EFFICIENCY OF THE NEW METHOD WAS FOUND TO BE QUITE HIGH. RELATED REPORTS ARE ED 003 044, ED 003 045, AND ED 003 046. (GD)

(5-0224) NAA

ED010607

STATISTICAL MODELS FOR THE EVALUATION AND INTERPRETATION OF EDUCATIONAL CRITERIA

Cooperative Research Project Number 1132

by

J. G. Saw and R. E. Flora

Virginia Polytechnic Institute
Blacksburg, Virginia
1965

PART II. THE PROBLEM OF CLASSIFYING MEMBERS OF A SINGLE POPULATION INTO GROUPS.

The presentation of these notes was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

ACKNOWLEDGEMENTS

The authors would like to thank the U. S. Department of Health Education and Welfare for the opportunity to investigate this particular problem and for their financial support. Thanks are extended to the faculty of the Statistics Department, Virginia Polytechnic Institute, for their interest and to Mrs. Ann Harrison for her excellent job of technical typing.

INDEX

CHAPTER 1

DATA PRESENTED BY GROUPS. DISCRIMINATION

1.1 Introduction

Frequently data is presented or collected by groups in an attempt to comment on possible differences between the groups. For example, the performance in a battery of tests may be tabulated for male and female; married and single; in-state or out-of-state students and a further classification may be according as the student was successful (obtained a B.S. for instance) or was unsuccessful (in this respect).

Whilst it is likely true that observations made within the in-state, male, married category; or the in-state, female, single category are approximately normally distributed, it is unlikely indeed that tests relating to mental performance are normally distributed within the class of people who obtain a bachelors degree. This fact is recognized quite easily by university teachers. Freshmens' performance in, say, mathematics is distributed quite close to normality; due to drop-outs over the following years some of the weaker students are not present in the senior year resulting in a discernable skewness in the grades, the better students being present in larger numbers. The grades of students accepted for graduate work are extremely skew, and in advanced graduate courses a final grade of A usually dominates all other grades; in fact, it is rare indeed to find a conscientious student in an advanced graduate course who does not achieve a grade of B or better. It would certainly be unwise therefore to work with data

relating to mental performance of graduate students under the assumption that these data were normally distributed.

It seems reasonable to assume that data collected on a student entering college would have a multivariate normal distribution (in the United States but possibly not in, for example, many European countries where students have to meet high qualifications in order to be permitted to embark on a higher education). Unfortunately, data pertaining to the weaker students is unlikely to be available in say the senior year due to the possibility of being a drop-out. However, a student may often be categorized in a broad way; for example it is possible to adopt the system of categorization

(i) student successful enough to proceed to a higher degree

(ii) student successful at the bachelors level but not likely to succeed in graduate school

(iii) students weak in their work at the bachelors level but who were not required to repeat or dropout

(iv) students who dropped out.

For every student then we have a vector of correlated normal variates $\underline{x}$ and an "index" d such that d takes on certain values according as the student belongs to group (i), (ii), (iii), or (iv) above. This should be contrasted with Part IV of this contract wherein each broad category contained only one student or vector observation and d took on the set of integers 1,2,3,...,n where n was the number of students sampled.

The classical approach to this problem has been to assume that $\underline{x}|d$ has a multivariate normal density, that is observations within a

broad category are normally distributed. This has been discussed in the previous paragraphs of this chapter and we reject the premise as unlikely.

Before presenting an analysis of the problem we define our data and notation. Suppose a sample of n individuals are available. Their performance on a battery of p tests is noted and is represented by the p elements of vector $\underline{x}$ ($\underline{x}' = (x_1, x_2, \ldots, x_p)$). In addition each student is categorized as described earlier. We may conveniently assign a characteristic random variable d to this student where d may take any of c different (scalar) value d_1, d_2, ..., etc. Our problem is to find: -

(a) which elements of $\underline{x}$ have the same distribution in each broad category

(b) which elements of $\underline{x}$ differ strikingly in different broad categories

(c) what linear combination of the elements of $\underline{x}$ are most indicative of the category to which the student is assigned.

We will list the vector observations for the n students as $\underline{x}_{11}, \underline{x}_{12}, \ldots, x_{1k_1}; \underline{x}_{21}, \underline{x}_{22}, \ldots, \underline{x}_{2k_2}; \ldots; \underline{x}_{c1}, \underline{x}_{c2}, \ldots, \underline{x}_{ck_c}$ where k_i students are assigned to category i and $k_1+k_2+\ldots k_c=n$. Sometimes the set $\{k_1, k_2, \ldots, k_i\}$ are themselves random variables; in some circumstances they may be fixed, for example in the case where we are concerned with the best three students we may set $c=2$ $k_1=n-3$; $k_2=3$.

1.2 The Classical Approach

In the classical approach it is assumed that

$$(1.2.1) \quad \underline{x}_{ij} \frown N_p(\underline{\mu}_i; V) \qquad j=1,2,\ldots,k_i;\ i=1,2,\ldots,c$$

that is $\underline{x}_{ij}$ has a multivariate normal density with mean $\underline{\mu}_i$ (possibly different for each category) but with a dispersion matrix which is the same (V) for all categories. We have envisaged situations wherein the assumptions of normality are invalid; the assumption of equal dispersion matrix is of course seldom ever valid.

It is required to find a system of multipliers, $\underline{\beta}$, such that the scalar quantity

$$(1.2.2) \quad Z_{ij} = \underline{\beta}'\underline{x}_{ij}$$

takes on distinctive values as i varies. Now if (1.2.1) holds then for fixed $\underline{\beta}$

$$(1.2.3) \quad Z_{ij} \frown N_1(\underline{\beta}'\underline{\mu}_i\ \underline{\beta}V\underline{\beta}) \qquad \begin{array}{l} j=1,2,\ldots \\ i=1,2,\ldots,c. \end{array}$$

so that the problem is that of selecting a $\underline{\beta}$ such that $\underline{\beta}'\underline{\mu}_i$ differ as widely as possible. Of course the $\{\underline{\mu}_i\}_{i=1}^{i=c}$ are unknown (otherwise there would be no problem) and the common dispersion matrix V is assumed unknown too.

Write

$$(1.2.4) \quad \begin{cases} \xi_i = \underline{\beta}'\underline{\mu}_i \\ \sigma^2 = \underline{\beta}'V\underline{\beta} \end{cases}$$

then in a likelihood test of the hypothesis: $\xi_1=\xi_2=\ldots=\xi_c$ we would construct for any given, fixed $\underline{\beta}$

$$F(\beta)=\frac{h/c-1}{e/n-c} \qquad (1.2.5)$$

where

$$h=\sum_{i=1}^{c}(\bar{Z}_i-\bar{Z}..)^2k_i \qquad (1.2.6)$$

$$e=\sum_{ij}\sum(Z_{ij}-\bar{Z}_i.)^2 \qquad (1.2.7)$$

Clearly $F(\underline{\beta})$ will take on small or large values according to the value of $\underline{\beta}$ since if $c<p$ we could find (if the $\{\underline{\mu}_i\}_{i=1}^{i=c}$ were known) a vector $\underline{\beta}$ such that $\xi_i=\underline{\beta}'\underline{\mu}_i=0$ for all $i=1,2,\ldots, c$. Alternatively, it would be possible to select a $\underline{\beta}$ such that $\xi_1<\xi_2<\ldots<\xi_c$ whenever $\underline{\mu}_1\neq\mu_2\neq\mu_3\neq\ldots\neq\mu_c$.

Classically one chooses the vector $\underline{\beta}$ which maximized the value of the ratio h/e.

Now

$$h=\sum_{i=1}^{c}(Z_i-Z..)^2k_i = \underline{\beta}'\{\Sigma(\underline{\bar{x}}_i-\underline{\bar{x}}..)(\underline{\bar{x}}_i-\bar{x}..)'k_i\}\underline{\beta} \qquad (1.2.8)$$

$$=\underline{\beta}'H\underline{\beta} \quad \text{(say)}$$

$$e=\sum_{ij}\sum(Z_{ij}-\bar{Z}_i.)^2=\underline{\beta}'\{\sum_{ij}\sum(\underline{x}_{ij}-\underline{\bar{x}}_i.)(\underline{x}_{ij}-\underline{\bar{x}}_i)'\}\underline{\beta} \qquad (1.2.9)$$

$$=\underline{\beta}'E\underline{\beta} \quad \text{(say).}$$

We therefore seek that $\underline{\beta}$ which maximizes

$$(1.2.10) \qquad F(\underline{\beta}) = \frac{\underline{\beta}'H\underline{\beta}}{\underline{\beta}'E\underline{\beta}}$$

where

$$(1.2.11) \qquad H=\sum_i (\overline{\underline{x}}_{i\cdot}-\overline{\underline{x}}_{\cdot\cdot})(\overline{\underline{x}}_{i\cdot}-\overline{\underline{x}}_{\cdot\cdot})'k_i$$

$$E=\sum_i\sum_j (\underline{x}_{ij}-\overline{\underline{x}}_{i\cdot})(\underline{x}_{ij}-\overline{\underline{x}}_{i\cdot})'$$

so that H and E are symmetric pxp matrices with H singular whenever $c<p$. We will hold $n>p>c$ so that E will be non-singular and H singular.

It follows quickly that $\hat{\underline{\beta}}$, the required value of $\underline{\beta}$, is given by

$$(1.2.12) \qquad (H-\theta E)\hat{\underline{\beta}} = \underline{0}$$

where θ is the largest root of

$$(1.2.13) \qquad |H-\theta E|=0$$

To emphasise we repeat the assumption made to get to this result.

Assumption I:

All vectors $\underline{x}_{ij}$ have a multivariate distribution.

Assumption II:

Each vector $\underline{x}_{ij}$ has the same dispersion matrix.

and also we repeat that if $\underline{x}_{ij}$ represent measures of intelligence and the categories are defined by differences in intelligence then neither assumption is likely to be valid.

1.3 An alternative approach

We now make less restrictive assumptions to meet the case where the mode of categorization is heavily related to the observed variates as described in the last paragraph of section 1.2.

We assume that if an individual is randomly drawn from some population then prior to categorization, $\underline{x}$, the vector of observations has a multivariate density. We shall not for the moment assume a specific type of multivariate density, only that a vector of means $\underline{\mu}$ and a dispersion matrix V exists for this density. The individual is now categorized and a value, d, characteristic of the category is assigned. Our n observations can now be represented as

$$\{d_i;\ \underline{x}_{ij}\} \qquad j=1,2,\ldots,k_i;\ i=1,2,\ldots,c.$$

We now seek a vector $\underline{\beta}$ and a set of values $d_1, d_2, \ldots d_c$ such that the correlation between the scalar quantitites d_i and $\underline{\beta}'\underline{x}_{ij}$ has as large a value as possible when taken over the n individuals. Let $\{\hat{d}_i\}_{i=1}^{i=c}$ and $\hat{\underline{\beta}}$ be the required system then

$$(1.3.1) \quad \frac{\sum\limits_{ij}\sum \hat{d}_i \hat{\underline{\beta}}'\underline{x}_{ij}}{\sqrt{\hat{\underline{\beta}}'\sum\limits_{ij}\sum(\underline{x}_{ij}-\underline{\bar{x}}_{\cdot\cdot})(\underline{x}_{ij}-\underline{\bar{x}}_{\cdot\cdot})'\hat{\underline{\beta}}}} > \frac{\sum\limits_{ij}\sum d_i \underline{\beta}'\underline{x}_{ij}}{\sqrt{\underline{\beta}'\sum\limits_{ij}\sum(\underline{x}_{ij}-\underline{\bar{x}}_{\cdot\cdot})(\underline{x}_{ij}-\underline{\bar{x}}_{\cdot\cdot})'\underline{\beta}}}$$

whenever $(\{d_i\},\underline{\beta}) \neq (\{\hat{d}_i\},\hat{\beta})$

It is convenient to "standardize" the $\{d_i\}$ by requiring

$$(1.3.2)\quad \begin{cases} \sum_i d_i k_i = 0. \\ \sum_i d_i^2 k_i = 1. \end{cases}$$

this leads to simplifications in the expressions for the correlations and also in the further development.

The numerator of the right-hand side of (1.3.1) can be written

$$\sum_i d_i \underline{\bar{x}}_{i.} k_i = \sum_i d_i k_i (\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..})$$

by virtue of (1.3.2).

Writing $T = \sum_{ij}\sum (\underline{x}_{ij} - \underline{\bar{x}}_{..})(\underline{x}_{ij} - \underline{\bar{x}}_{..})'$

the denominator is then $\sqrt{\underline{\beta}' T \underline{\beta}}$.

Our objective is to maximize corr $(\{d_i\}, \underline{\beta})$ where

$$(1.3.3)\quad \operatorname{corr}(\{d_i\}, \underline{\beta}) = \frac{\underline{\beta}' \sum_i d_i k_i (\underline{\bar{x}}_i - \underline{\bar{x}}_{..})}{\sqrt{\underline{\beta}' T \underline{\beta}}}$$

subject to $\sum_i d_i k_i = 0$; $\sum_i \sum_i d_i^2 k_i = 1$ over choice of $\{d_i\}_{i=1}^{i=c}$ and $\underline{\beta}$.

Since the scale of $\underline{\beta}$ is immaterial we choose to maximize corr$(\{d_i\}, \underline{\beta})$ subject to $\underline{\beta}' T \underline{\beta} = 1$.

Case c=2

In the case c=2, the conditions $\sum_{i=1}^{2} d_i k_i = 0$ and $\sum_{i=1}^{2} d_i^2 k_i = 1$ are sufficient to determine d_1 and d_2 since we have

$$(1.3.4)\quad \begin{cases} d_1k_1+d_2k_2=0 \\ d_1^2k_1+d_2^2k_2=1 \end{cases} \qquad k_1+k_2=n.$$

it follows that

$$(1.3.5)\quad \begin{cases} \hat{d}_1= -\sqrt{k_2/k_1n} \\ \hat{d}_2= \sqrt{k_1/k_2n}. \end{cases}$$

In this event

$$(1.3.6)\quad \{\mathrm{corr}(d_1,d_2,\underline{\beta})\}^2=\frac{\underline{\beta}'(\sum_i d_i^2k_i^2(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.}-\underline{\bar{x}}_{..})')\underline{\beta}}{\underline{\beta}'T\underline{\beta}}$$

$$=\frac{\frac{k_1k_2}{n}\ \underline{\beta}'\sum_{i=1}^{2}(\underline{\bar{x}}_i-\underline{\bar{x}}_{..})(\underline{x}_i-\underline{\bar{x}}_{..})'\underline{\beta}}{\underline{\beta}'T\underline{\beta}}$$

$$=\frac{\frac{k_1k_2}{n}\ \underline{\beta}'(\underline{\bar{x}}_{1.}-\underline{\bar{x}}_{2.})(\underline{\bar{x}}_{1.}-\underline{\bar{x}}_{2})'\underline{\beta}}{\underline{\beta}'T\underline{\beta}}$$

and thus $\underline{\beta}$ is the solution of

$$(1.3.8)\quad \left(\frac{k_1k_2}{n}(\underline{\bar{x}}_{1.}-\underline{\bar{x}}_{2.})(\underline{\bar{x}}_{1.}-\underline{\bar{x}}_{2.})'-\lambda T\right)\hat{\underline{\beta}}=0$$

where λ is the largest root of

$$(1.3.9)\quad \left|\frac{k_1k_2}{n}(\underline{\bar{x}}_{1.}-\underline{\bar{x}}_{2.})(\underline{\bar{x}}_{1.}-\underline{\bar{x}}_{2})'-\lambda T\right|=0.$$

Now since $(\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)(\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)'$ is of rank 1 there is only one non-zero λ which is a solution of (1.3.9).

It is noted that

(1.3.10) $$\frac{k_1k_2}{n}(\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)(\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)'=\sum_i(\underline{\bar{x}}_i.-\underline{\bar{x}}..)(\underline{\bar{x}}_i.-\underline{\bar{x}}..)'k_i$$

$$=H$$

where H corresponds to the matrix H defined in the classical situation (equation (1.2.8)). Also

(1.3.11) $$T=E+H$$

where E is defined in 1.2.9. Hence λ is the largest root of

(1.3.12) $$|H-\lambda(E+H)|=0$$

and $\hat{\underline{\beta}}$ the associated eigen vector.

Evidently with c=2 and θ defined by (1.2.13) then

(1.3.13) $$\lambda=\frac{\theta}{1+\theta}\ .$$

Now if

(1.3.14) $$\left(H-\frac{\theta}{1+\theta}(E+H)\right)\hat{\underline{\beta}}=0.$$

then

(1.3.15) $$0=\left((1+\theta)H-\theta(E+H)\right)\hat{\underline{\beta}}$$

$$=(H-\theta E)\hat{\underline{\beta}}$$

Since λ is an increasing function of θ we see that the vector $\underline{\beta}$ is the same in our new approach as for the classical approach (c=2), a somewhat startling result.

Case c>2.

In the case of more than two categories we shall maximize $(\mathrm{corr}\ (\{d_i\},\underline{\beta}))^2$ defined by (1.3.3) over choice of $\{d_i\}$ satisfying (1.3.2) holding the scale of $\underline{\beta}$ so that $\beta'T\underline{\beta}=1$.

Let ϕ_1,ϕ_2 and λ be undetermined multipliers and construct

$$Q=\left(\underline{\beta}'\sum_i d_i k_i (\underline{\bar{x}}_i-\underline{\bar{x}}..)\right)^2-2\phi_1\Sigma d_i k_i-\phi_2(\Sigma d_i^2 k_i-1)-\lambda(\underline{\beta}'T\underline{\beta}-1)$$

Using the theory of Lagrange undertermined multipliers the required solution is given by

$$\frac{\partial Q}{\partial d_i}=0;\ \frac{\partial Q}{\partial \underline{\beta}}=\underline{0};\ \frac{\partial Q}{\partial \phi_1}=\frac{\partial Q}{\partial \phi_2}=\frac{\partial Q}{\partial \lambda}=0.$$

The maximal value is achieved as $\{\hat{d}_i\},\hat{\underline{\beta}}$ so that

(1.3.16) $$\left(\hat{\underline{\beta}}'\Sigma d_i k_i (\underline{\bar{x}}_i.-\underline{\bar{x}}..)\right)k_i(\underline{\bar{x}}_i-\underline{\bar{x}}..)'\hat{\underline{\beta}}=\phi_1 k_i+\phi_2 d_i k_i$$

$i=1,2,\ldots c.$

(1.3.17) $$\left(\Sigma\hat{d}_i k_i(\underline{\bar{x}}_i.-\underline{\bar{x}}..)\Sigma d_i k_i(\underline{\bar{x}}_i.-\underline{\bar{x}}..)'-\lambda T\right)\hat{\underline{\beta}}=0$$

(1.3.18) $$\begin{cases}\Sigma\hat{d}_i k_i=0\\ \Sigma\hat{d}_i{}^2 k_i=1\\ \hat{\underline{\beta}}'T\hat{\underline{\beta}}=1\end{cases}$$

Summing (1.2.17) over i we have $\phi_1=0$ whence

(1.3.19) $$\hat{d}_i\ \alpha\ (\underline{\bar{x}}_i.-\underline{\bar{x}}..)\hat{\underline{\beta}}$$

Using (1.3.18) it is easily established that

$$(1.3.20)\quad \hat{d}_i = (\underline{\bar{x}}_i - \underline{\bar{x}}_{..})'\hat{\underline{\beta}} \Big/ \sqrt{\underline{\beta}'H\underline{\beta}}$$

where, as usual,

$$(1.3.21)\quad H = \sum_i (\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..})'k_i$$

Premultiplying (1.2.18) by $\hat{\underline{\beta}}'$ we find

$$(1.3.22)\quad \lambda = \max_{\{d_i\},\underline{\beta}} (\mathrm{corr}(\{d_i\},\underline{\beta}))^2$$

But

$$(1.3.23)\quad \Sigma d_i k_i (\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..}) \Sigma d_i k_i (\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..})'$$

$$= \frac{H\hat{\underline{\beta}}\hat{\underline{\beta}}'H}{\hat{\underline{\beta}}'H\hat{\underline{\beta}}}$$

therefore λ is the largest solution of

$$(1.3.24)\quad \frac{H\hat{\underline{\beta}}\hat{\underline{\beta}}'H\hat{\underline{\beta}}}{\hat{\underline{\beta}}'H\hat{\underline{\beta}}} = \lambda T\hat{\underline{\beta}} \quad ,$$

that is, the largest solution of

$$(1.3.25)\quad (H - \lambda T)\hat{\underline{\beta}} = 0$$

Define E by E=T-H so that, as in the classical case,

$$(1.3.26)\quad E = \sum_i \sum_j (\underline{x}_{ij} - \underline{\bar{x}}_{i.})(\underline{x}_{ij} - \underline{\bar{x}}_{i.})'$$

then

λ is the largest root of $|H-\lambda(E+H)|=0$ and $\hat{\underline{\beta}}$ the associated eigen vector. Alternatively, with $\lambda=\theta/(1+\theta)$, θ is the largest root of $|H-\theta E|=0$ and $\hat{\underline{\beta}}$ satisfies $(H-\theta E)\hat{\underline{\beta}}=\underline{0}$. Either way, $\hat{d}_i$ is proportional to $(\underline{\bar{x}}_{i\cdot}-\underline{\bar{x}}_{\cdot\cdot})'\hat{\underline{\beta}}$.

It is seen then that for any c, the discriminant is numerically the same for our system as in the classical case despite the fact that our basic assumptions are widely different.

1.4 An example with three broad categories.

A group of eighteen math majors are critically examined by their faculty with the purpose of evaluating the departmental admissions examination. The faculty agreed to divide the eighteen into three groups of six according to the performance of the students in their junior and senior years. After this has been done, the respective performances in the four-part admissions exam were collated with the students. The results are given below.

Top group: $\begin{bmatrix}88\\87\\92\\88\end{bmatrix}$ $\begin{bmatrix}82\\93\\98\\91\end{bmatrix}$ $\begin{bmatrix}84\\94\\100\\84\end{bmatrix}$ $\begin{bmatrix}70\\94\\82\\82\end{bmatrix}$ $\begin{bmatrix}88\\90\\89\\71\end{bmatrix}$ $\begin{bmatrix}90\\88\\93\\88\end{bmatrix}$

Middle group: $\begin{bmatrix}72\\94\\72\\81\end{bmatrix}$ $\begin{bmatrix}100\\70\\94\\90\end{bmatrix}$ $\begin{bmatrix}86\\88\\86\\73\end{bmatrix}$ $\begin{bmatrix}94\\79\\85\\76\end{bmatrix}$ $\begin{bmatrix}97\\68\\90\\93\end{bmatrix}$ $\begin{bmatrix}84\\76\\74\\71\end{bmatrix}$

Low group: $\begin{bmatrix}80\\71\\71\\80\end{bmatrix}$ $\begin{bmatrix}70\\77\\69\\83\end{bmatrix}$ $\begin{bmatrix}89\\77\\69\\65\end{bmatrix}$ $\begin{bmatrix}80\\69\\69\\80\end{bmatrix}$ $\begin{bmatrix}69\\76\\68\\84\end{bmatrix}$ $\begin{bmatrix}91\\76\\71\\66\end{bmatrix}$

The vector is

$$\begin{bmatrix}\text{score in part I of admissions exam}\\ \text{" " " II " " "}\\ \text{" " " III " " "}\\ \text{" " " IV " " "}\end{bmatrix}$$

It is the admissions examination which the faculty require to evaluate with a view to using it to reject students unlikely to derive great benefit from a college education. In broad terms one may ask: "is the admission exam any indication of performance in college; if so how should the results or scores on that examination be used to select the more promising students."

Although we have not yet developed the distribution theory relevant to this problem we have available a technique indicated: the previous section. It is believed that, taken over the population of students, each admission score if is ! approximately normally distributed and is not a mixture of populations as is required by the theory of discriminant functions.

Using the notation of the previous section we have.

$k_1=k_2=k_3=6$.

$$\underline{\bar{x}}_{1.} = \frac{1}{6}\begin{bmatrix}502\\546\\554\\504\end{bmatrix} \qquad \underline{\bar{x}}_{2.} = \frac{1}{6}\begin{bmatrix}533\\475\\501\\484\end{bmatrix} \qquad \underline{\bar{x}}_{3.} = \frac{1}{6}\begin{bmatrix}479\\446\\417\\458\end{bmatrix}$$

$$= \qquad \underline{\bar{x}}_{..} = \frac{1}{18}\begin{bmatrix}1514\\1467\\1472\\1446\end{bmatrix}$$

$$H = \sum_i (\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..})(\underline{\bar{x}}_{i.} - \underline{\bar{x}}_{..})' k_i$$

$$= \frac{1}{18}\begin{bmatrix}4406 & 1665 & 6044 & 1842\\1665 & 15882 & 19899 & 6774\\6044 & 19899 & 28634 & 9546\\1842 & 6774 & 9546 & 3192\end{bmatrix}$$

$$T = \sum_{ij}\sum (\underline{x}_{ij} - \underline{\bar{x}}_{..})(\underline{x}_{ij} - \underline{\bar{x}}_{..})'$$

$$= \frac{1}{12}\begin{bmatrix}26420 & -8064 & 16244 & -336\\-8064 & 27117 & 14742 & 1188\\16244 & 14742 & 39440 & 15378\\-336 & 1188 & 15378 & 21780\end{bmatrix}$$

$$E = T - H = \sum_{ij}\sum (\underline{x}_{ij} - \underline{\bar{x}}_{i.})(\underline{x}_{ij} - \underline{\bar{x}}_{i.})'$$

$$= \frac{1}{12}\begin{bmatrix}22014 & -9729 & 10200 & -2178\\-9729 & 11235 & -5157 & -5586\\10200 & -5157 & 10806 & +5832\\-2178 & -5586 & 5832 & 18588\end{bmatrix}$$

We need now to extract the largest root of

$$|H-\lambda(E+H)|=0$$

(see equation (1.3.12)).

Since $E+H=T$ is positive definite we can find a non-singular matrix M such that $MTM'=I$. This matrix is of course obtained by performing a Doolittle on T. With F the Doolittle lower diagonal matrix with $\{f_{ii}=1\}$ then

$$FTF'=D \qquad \text{with } \left.\begin{array}{ll} d_{ij}=0 & i\neq j \\ d_{ii}>0. & \end{array}\right\}$$

Whence $M=D^{-\frac{1}{2}}F$. We now require the largest latent root of MHM' which can be obtained by the itterative methods. It is noted that in the case of two groups, H, and therefore MHM', is of rank one whereupon largest latent root of MHM' = trace MHM'. In our case, the largest latent root is

$$\lambda = 0.8802$$

and the associated vector is $\begin{bmatrix} 0.1165 \\ 10.5208 \\ 11.1623 \\ 1.0000 \end{bmatrix}$

or any multiple of it.

The $\{\hat{d}_i\}$ are now found from equation 1.2.20 to be (after standardization).

(0.2943; 0.00765; -0.28477).

It is of interest to obtain $\underline{x}'\underline{\hat{\beta}}$ for each of the 18 students:-

Top group:	2040	2173	2198	1994	2021	2062
Middle group:	1882	1887	1969	1867	1824	1706
Low group:	1628	1671	1655	1585	1650	1669

The $\underline{\beta}$ used was $\begin{bmatrix} 0.1165 \\ 10.5208 \\ 11.1625 \\ 1.0000 \end{bmatrix}$.

The resulting values $\underline{x}'\underline{\beta}$ divide themselves very nicely as we would expect since there is to be quite a large (0.8802) correlation between the $\underline{x}'\underline{\hat{\beta}}$ and the $\{d_i\}$.

Given any other student (not a member of the set of eighteen) and his admissions exam score $\underline{X}$ we could compute $\underline{X}'\underline{\hat{\beta}}=\underline{Z}$ to decide which category this new individual is most likely to belong to.

Clearly the admissions exam is of value though only the part II and part III scores appear to have a great deal of correlation with the final group to which a student is judged to belong.

Some common sense has to be applied to provide group boundaries for the Top, Middle and Low groups. Looking at the 18 values of $\underline{x}'\underline{\hat{\beta}}$, a suitable system might be.

judge as top group if $\underline{X}'\underline{\beta}>2000$

judge as middle group if $1700<\underline{X}'\underline{\beta}<2000$

judge as low group if $X'\underline{\hat{\beta}}\leq 1700$.

The probability of incorrect assignment of course depends on the actual value of $\underline{X}'\underline{\hat{\beta}}$ (not true in the classical case). Obviously if for some student $\underline{X}'\underline{\hat{\beta}}=2400$ then we would feel pretty safe in judging him to be in the top group.

Comparisons can also be made between new students. Suppose $\underline{X}_A$ and $\underline{X}_B$ are the admissions scores of two new students A and B then we may judge that A will succeed better than B if

$$\underline{X}_A'\underline{\hat{\beta}} > \underline{X}_B'\underline{\hat{\beta}}$$

again, the confidence of our judgement will be a function of

$$(\underline{X}_A'-\underline{X}_B')\underline{\hat{\beta}}.$$

CHAPTER 2

NULL DISTRIBUTION

2.1 The null distribution of the largest latent root.

From equation (1.3.26) we find that the values of $\underline{\hat{\beta}}$ and $\{\hat{d}_i\}$ hinge on the value of the largest latent root, λ, of H in the metric of H+E; that is, with

$$H = \sum_i (\underline{\bar{x}}_{i\cdot}-\underline{\bar{x}}_{\cdot\cdot})(\underline{\bar{x}}_{i\cdot}-\underline{\bar{x}}_{\cdot\cdot})'k_i$$

$$E = \sum_{ij}\sum (\underline{x}_{ij}-\underline{\bar{x}}_{i\cdot})(\underline{x}_{ij}-\underline{\bar{x}}_{i\cdot})'$$

then λ is the largest solution of

$$|H-\lambda(H+E)|=0.$$

If no element of $\underline{x}$ is correlated with the grouping scheme, that is

$$p(\underline{x}|\text{ individual categorized in group } j)$$

is independent of j then the grouping of the individuals results in a random arrangement of the $\underline{x}_{ij}$. We have then that the density of $\underline{x}_{ij}$ is independent of j and also of i. It is then easy to show that H and E are independent variates of the Wishart type when $\underline{x}_{ij}$ is assumed to have a multivariate normal density; in fact

$$E \sim \mathcal{W}_p(V;n-c:(0))$$

$$H \sim \mathcal{W}_p(V;c-1:(0))$$

whenever

$\underline{x}_{ij} \frown N_p(\underline{\mu}:V)$

[for notation see Part 4, volume I of this contract].

The joint distribution of the latent roots of $|H-\phi(E+H)|$ is known, see for example Part 4, volume I, chapter 6 of this contract, and good approximations to the percentage points of the largest root are available. The case where c=2 can be be handled in terms of the percentage points of the F-ratio.

2.2 Case c=2.

When c=2 we find that H is of rank 1, consequently there exists only one non-zero solution of equation (1.2.26). Returning to section 1.3 we see that

$$(2.2.1) \quad H=\frac{k_1k_2}{n}(\underline{\bar{x}}_{1\cdot}-\underline{\bar{x}}_{2\cdot})(\underline{\bar{x}}_{1\cdot}-\underline{\bar{x}}_{2\cdot})'$$

Writing $\theta=\lambda/1-\lambda$, then θ is the non-zero solution of

$$(2.2.2) \quad \left|\frac{k_1k_2}{k_1+k_2}(\underline{\bar{x}}_{1\cdot}-\underline{\bar{x}}_{2\cdot})(\underline{\bar{x}}_{1\cdot}-\underline{\bar{x}}_{2\cdot})'-\theta E\right| = 0$$

Now since for $\theta\neq 0$

$$|\theta E-\underline{u}\underline{u}'| = \left|\begin{array}{c|c} 1 & \underline{u}' \\ \hline \underline{u} & \theta E \end{array}\right| = |\theta E|\left(1-\frac{Z'E^{-1}Z}{\theta}\right)$$

we see that

$$(2.2.3) \quad \theta = \frac{\lambda}{1-\lambda} = \frac{k_1 k_2}{k_1+k_2} (\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)' E^{-1} (\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)$$

which, given E and $(\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)$, is easy to calculate by performing a Doolittle on E carrying $(\underline{\bar{x}}_1.-\underline{\bar{x}}_2.)$ [Part 4 Volume II of this contract]. Finally

$$(2.2.4) \quad \theta \times \frac{n-p-1}{p} \Big| \; Ho \sim F_{p:n-p-1}$$

large values of θ are significant.

2.3 Case $c>2$.

In the case $c>2$ we are committed to the use of percentage points of the largest λ satisfying

$$|H-\lambda(H+E)|=0$$

the distribution of the largest latent root is discussed in the statistical literature by several authors. In particular, tables of percentage points are given by D.L. Heck(Ann. Math. Statist., 1960, pp. 625-642).

As an alternative test of the hypothesis Ro.(1,2,..p)=0 one may also use

$$-2\rho_o \log|E|/|E+H| \;\hat{\sim}\; \chi^2_{p(c-1)}$$

where $\rho_o=n-1-\frac{1}{2}(c+p)$.

This does not require the evaluation of the largest root and is computationally much simplier, however it is to be remembered that if $-2\rho_o \log|E|/|E+H|$ proves significant, the next step will likely be to establish $\hat{\underline{\beta}}$ for which the largest latent root will be needed.

CHAPTER 3.

THE ALTERNATIVE DISTRIBUTION

Case c=2

3.1 Preliminaries

For the case c=2, the multiple correlation coefficient, R say, between the vectors $\underline{x}_{ij}$ $j=1,2,..k_i$; $i=1,2$ and the scores (see equation 1.3.5)

$$(3.1.1)\quad \begin{cases} d_1 = -\sqrt{k_2/k_1 n} \\ d_2 = \sqrt{k_1/k_2 n} \end{cases}$$

may be obtained from

$$(3.1.2)\quad R^2 = \Big(\sum_{ij}\sum d_i(\underline{x}_{ij}-\bar{\underline{x}}..)\Big)C^{-1}\Big(\sum_{ij}\sum d_i(\underline{x}_{ij}-\bar{\underline{x}}..)\Big)'$$

$$= (\sum_i d_i k_i \bar{\underline{x}}_i.)C^{-1}(\sum_i d_i k_i \bar{\underline{x}}_i.)'$$

$$= \frac{k_1 k_2}{n}(\bar{\underline{x}}_1.\ \bar{\underline{x}}_2.)'C^{-1}(\bar{\underline{x}}_1.-\bar{\underline{x}}_2.)$$

where

$$(3.1.3)\quad C = \sum_{ij}\sum(\underline{x}_{ij}-\bar{\underline{x}}..)(\underline{x}_{ij}-\bar{\underline{x}}..)'$$

Accordingly, we may use some of the work developed in Part I of this contract (in particular chapter 2 of that part) for a suitable underlying model.

3.2 An underlying model.

Assume that an unknown variable x_{0ij} exists for every individual. For the n individuals whose vector measurement $\underline{x}_{ij}$ is available $(i=1, 2: j=1,2,..k_i)$ there is the variable x_{0ij} which is unobserved but about which the following information is available.

(3.2.1) $\sup_j\{x_{01j}\}<\inf_j\{x_{02j}\}$

that is the variable x_{01j_1} for an individual is group 1 is less than the x_{02j_2} for any individual in group 2. We may imagine then that the dichotomy of n individuals into the two groups is made on the basis of the unobservable x_{0ij}(hereafter called the characteristic normal variable).

It is recognized that the existence of the characteristic normal variable represents an idealized situation. In effect we are saying that the categorization of an individual is made on the basis of a single characteristic variable. Obviously one can think of many ways of assigning an individual to one of various categories; however the use of the characteristic normal variable seems to use not unreasonable and fairly simple to apply. Further the critical regions associated with the test of hypothesis H_o:R=0 do not depend upon the assumptions of the existence of characteristic normal variables although of course the investigation of the power of the critical region does.

We shall suppose

$$(3.2.2)\quad \begin{bmatrix} x_o \\ \underline{x} \end{bmatrix} \frown N_{p+1}\left(\begin{bmatrix} \mu_o \\ \underline{\mu} \end{bmatrix} ; \begin{bmatrix} v_{oo} & \underline{v}_{01}' \\ \underline{v}_{10} & V \end{bmatrix} \right)$$

where $\underline{x}$ is the observed vector and x_0 the corresponding characteristic normal variable. That is to say $(x_o; x_1, x_2 .. x_p)$ have a multivariate normal density with mean vector and dispersion matrix as displayed in (3.2.2).

If it be required to divide n individuals into two categories with k_1 and k_2 members then we construct two groups

$\{x_{011}, x_{012}, \ldots\ldots x_{01k_1}\}$ and $\{x_{021}\ x_{022} .. x_{02k_2}\}$

which passes the property given in (3.2.1) that is

$$\sup_j \{x_{01j}\} < \inf_j \{x_{02}\}.$$

the groups are therefore uniquely constructed with probability one.

It is noted that in another problem k_1 and k_2 $(k_1+k_2=n)$ maybe random variables; the dichotomy being achieved via

$$\sup_j \{x_{01j}\} < \tau < \inf_j \{x_{02j}\}).$$

<u>3.3 Development</u>.

The development of the distribution of R^2 now parallels

the development in Part I of this contract. In particular we have

$$(3.3.1)\quad \mathcal{E}\{1-R^2\}^h = \mathcal{E}\left\{\frac{w_{11}}{w_{11}+h_{11}}\right\}^h \frac{\left(\frac{v-p}{2}+h-1\right)^{[h]}}{\left(\frac{v-1}{2}+h-1\right)^{[h]}}$$

$(v=n-1)$

where

$$(3.3.2)\quad a^{[r]} = a(a-1)(a-2)\ldots(a-r+1)$$

$$(3.3.3)\quad h_{11} \mid \{x_{0ij}\} \sim \chi_1^2(\lambda_1)$$

$$(3.3.4)\quad w_{11} \mid \{x_{0ij}\} \sim \chi_{n-2}^2(\lambda_2).$$

w_{11} and h_{11} are independently distributed conditional on $\{x_{0ij}\}$.

$$(3.3.5)\quad \lambda_1 = \delta^2\left(\sum_{ij}\sum d_i x_{0ij}\right)^2$$

$$(3.3.6)\quad \lambda_1+\lambda_2 = \delta^2 \sum_{ij}\sum (x_{0ij}-\bar{x}_{0\cdot\cdot})^2$$

with

$$(3.3.7)\quad \delta = R_0/\sqrt{1-R_0^2}$$

and R_0 the population multiple correlation coefficient between x_0 and $(x_1, x_2, \ldots, x_p)$.

Upon substituting values for d_1 and d_2

$$(3.3.8)\quad \lambda_1 = \tfrac{1}{2}\delta^2 \frac{k_1 k_2}{n} (\bar{x}_{02}-\bar{x}_{01})^2$$

Writing

$$(3.3.9)\quad s^2=\sum_{ij}\sum(x_{0ij}-\bar{x}_{0}..)^2$$

$$(3.3.10)\quad \eta^2=\frac{k_1k_2}{n}(\bar{x}_{02}-\bar{x}_{01})^2$$

we have

$$(3.3.11)\quad \begin{cases}\lambda_1=\delta^2 s^2\\ \lambda_2=\delta^2(s^2-\eta^2)\end{cases}$$

From the form of equation 3.3.1 evidently

$$(3.3.12)\quad \mathcal{E}(1-R^2)^h=\mathcal{E}(1-r^2)^h\frac{\left(\frac{v-p}{2}+h-1\right)^{[h]}}{\left(\frac{v-1}{2}+h-1\right)^{[h]}}$$

where r is the "multiple correlation" in the case p=1 and r=n-1.

3.4 The case p=1; moments of a correlation coefficient.

In the last section, we saw that the moments of our test criterion (general p) were a simple function of the moments of a test criterion appropriate to the case p=1. Write

$$(3.4.1)\quad r=\sum_{ij}\sum d_i x_{ij}\Big/\sum_{ij}\sum(x_{ij}-\bar{x}..)^2$$

and suppose

$$(3.4.2)\quad x_\alpha | x_{0\alpha} \frown N_1(\rho x_{0\alpha}: 1-\rho^2) \qquad \alpha=1,2,\ldots,n.$$

then with

$$(3.4.3)\quad \delta=\rho/\sqrt{1-\rho^2}$$

we find, using the techniques of chapter 3, part I of this contract that

$$(3.4.4)\quad 1-r^2 \frown w_{11}/(w_{11}+h_{11})$$

where w_{11} and h_{11} have independent χ^2-distribution conditional on $\{x_{0\alpha}\}_{\alpha=1}^{\alpha=n}$. In fact

$$(3.4.5)\quad h_{11}|\{x_{0ij}\} \frown \chi_1^2(\lambda_1)$$

$$(3.4.6)\quad w_{11}|\{x_{0ij}\} \frown \chi_\nu^2(\lambda_2)$$

with λ_1 and λ_2 defined as in (3.3.11), that is

$$(3.4.7)\quad \begin{cases} \lambda_1=\delta^2 S^2 \\ \lambda_2=\delta^2(S^2-\eta^2) \end{cases}$$

and

$$(3.4.8)\quad S^2=\sum_i\sum_j(x_{0ij}-\bar{x}_{0..})^2$$

$$(3.4.9)\quad \eta^2=\frac{k_1k_2}{n}\,(\bar{x}_{02}-\bar{x}_{01})^2$$

The condition joint density of h_{11} and w_{11} is

(3.4.10) $p(h_{11},w_{11}|\{x_{oij}\})=$

$$\frac{e^{-\frac{1}{2}(\lambda_1+\lambda_2)}\, e^{-\frac{1}{2}(h_{11}+w_{11})}}{2^{\frac{v+1}{2}}} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{h_{11}^{\frac{1}{2}+i-1}\, w_{11}^{\frac{v}{2}+j-1}\lambda_1^i\lambda_2^j}{2^{2i}i!\,\Gamma(\frac{1}{2}+i)\,2^{2j}j!\,\Gamma(\frac{1}{2}+j)}$$

Transforming from (h_{11},w_{11}) to (Z,w_{11}) where

(3.4.11) $Z=w_{11}/(w_{11}+h_{11})=1-r^2$

and integrating out the w_{11} we find

(3.4.12) $p(Z|\{x_{oij}\})=$

$$e^{-\frac{1}{2}(\lambda_1+\lambda_2)} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \frac{Z^{\frac{v}{2}+j-1}\,(1-Z)^{\frac{1}{2}+i-1}}{B(\frac{v}{2}+j;\frac{1}{2}+i)} \; \frac{\lambda_1^i\lambda_2^j}{i!\,j!\,2^{i+j}}$$

over the region $0\le Z\le 1$.

The moments of r^2 condition on $\{x_{oij}\}$ are now a simple matter. In fact

(3.4.13) $\mathcal{E}(r^2|\{x_{ioj}\})=$

$$e^{-\frac{1}{2}(\lambda_1+\lambda_2)}\left[\frac{1}{2}\sum_{k=0} \frac{(\lambda_1+\lambda_2)^k}{k!\,2^k(\frac{n-1}{2}+k)} + \frac{1}{2}\lambda_1\sum_{0}^{\infty}\frac{(\lambda_1+\lambda_2)^k}{k!\,2^k(\frac{n+1}{2}+k)}\right]$$

$$(3.4.14)\quad \mathcal{E}(r^4|\{x_{oij}\})\cdot e^{1/2(\lambda_1+\lambda_2)} =$$

$$\frac{3}{4}\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+1}{2}+k)}[2] + \frac{3}{2}\lambda_1\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+3}{2}+k)}[2]$$

$$+(\tfrac{1}{2}\lambda_1)^2\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+5}{2}+k)}[2]$$

$$(3.4.15)\quad \mathcal{E}(r^6|\{x_{0ij}\})\; e^{1/2(\lambda_1+\lambda_2)} =$$

$$\frac{15}{8}\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+3}{2}+k)}[3] + \frac{45}{8}\lambda_1\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+5}{2}+k)}[3]$$

$$+\frac{15}{2}(\tfrac{1}{2}\lambda_1)^2\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+7}{2}+k)}[3] + (\tfrac{1}{2}\lambda_1)^3\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+9}{2}+k)}[3]$$

$$(3.4.16)\quad \mathcal{E}(r^8|\{x_{oij}\})\; e^{-\frac{1}{2}(\lambda_1+\lambda_2)} =$$

$$\frac{105}{16}\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+5}{2}+k)}[4] + \frac{105}{4}\lambda_1\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+7}{2}+k)}[4]$$

$$\frac{105}{2}(\tfrac{1}{2}\lambda_1)^2\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+9}{2}+k)}[4] + 14(\tfrac{1}{2}\lambda_1)^3\sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!2^k(\frac{n+11}{2}+k)}[4]$$

$$+\left(\frac{1}{2}\lambda_1\right)^4 \sum_0^\infty \frac{(\lambda_1+\lambda_2)^k}{k!\,2^k\,\Gamma\left(\frac{n+13}{2}+k\right)} \quad [4]$$

To work toward the non-conditional moments we use the fact that, for $\{x_{0\alpha}\}_{\alpha=1}^{\alpha=n}$ a set of independent normal variates

$$(3.4.17) \quad p\left(\frac{\eta^2}{s^2},\ s^2\right) = p\left(\frac{\eta^2}{s^2}\right)p(s^2)$$

that is η^2/s^2 and s^2 are independently distributed. Further s^2 has a central χ^2-distribution with $n-1=v$ degrees of freedom.

Consequently

$$(3.4.18) \quad \mathcal{E}\{s^{2k}e^{-\frac{1}{2}\delta^2 s^2}\} = 2^k(1+\delta^2)^{\frac{n-1-2k}{2}} \cdot \frac{\Gamma\left(\frac{n-1}{2}+k\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

Now, using (3.4.17) and

$$(3.4.19) \quad \mathcal{E}(\eta^2/s^2)^h = \mathcal{E}(\eta^2)^h/\mathcal{E}(s^2)^h$$

we have

$$(3.4.20) \quad \mathcal{E}(\eta^{2h}s^{2k}e^{-\frac{1}{2}\delta^2 s^2})$$

$$= \mathcal{E}\left(\left(\frac{\eta^2}{s^2}\right)^h s^{2(h+k)} e^{-\frac{1}{2}\delta^2 s^2}\right)$$

$$= \mathcal{E}\left(\frac{\eta^2}{s^2}\right)^h \left(s^{2(h+k)}e^{-1/2\,\delta^2 s^2}\right)$$

$$= \mathcal{E}\left(\frac{n}{s^2}\right)^{2h} \frac{2^{k+h}}{(1+\delta^2)^{\frac{n-1}{2}+h+k}} \frac{\Gamma\left(\frac{n-1}{2}+k+h\right)}{\Gamma\left(\frac{n-1}{2}\right)}$$

$$= \frac{\mathcal{E}(n^{2h})\, 2^{k+h}\, \Gamma\left(\frac{n-1}{2}+k+h\right)}{(n-1)(n+1)\ldots(n-1+2h)\,\Gamma\left(\frac{n-1}{2}\right)(1+\delta^2)^{\left(\frac{n-1}{2}+h+k\right)}}$$

The nonconditional moments of r^2 are then, using (3.4.20)

(3.4.21) $\mathcal{E}_{\rho}(r^2) =$

$$\frac{1}{2}\sum_{0}^{\infty} \frac{\delta^{2k}}{(1+\delta^2)^{\left(\frac{n-1}{2}+k\right)}} \frac{\left(\frac{n-3}{2}+k\right)^{[k]}}{k!\left(\frac{n-1}{2}+k\right)}$$

$$+\frac{1}{2}\mathcal{E}(n^2)\sum_{0}^{\infty} \frac{\delta^{2k+2}}{(1+\delta^2)^{\left(\frac{n+1}{2}+k\right)}} \frac{\left(\frac{n-1}{2}+k\right)^{[k]}}{k!\left(\frac{n+1}{2}+k\right)}$$

(3.4.22) $\mathcal{E}(r^4) =$

$$\frac{3}{4}\sum_{0}^{\infty} \frac{\delta^{2k}}{(1+\delta^2)^{\left(\frac{n-1}{2}+k\right)}} \frac{\left(\frac{n-3}{2}+k\right)^{[k]}}{k!\left(\frac{n+1}{2}+k\right)^{[2]}}$$

$$+\frac{3}{2}\mathcal{E}(n^2)\sum_{0}^{\infty} \frac{\delta^{2k+2}}{(1+\delta^2)^{\left(\frac{n+1}{2}+k\right)}} \frac{\left(\frac{n-1}{2}+k\right)^{[k]}}{k!\left(\frac{n+3}{2}+k\right)^{[2]}}$$

$$\frac{1}{4}\mathcal{E}(\eta^4)\sum_{0}^{\infty}\frac{\delta^{2k+4}}{(1+\delta^2)^{\left(\frac{n+3}{2}+k\right)}}\quad\frac{\left(\frac{n+1}{2}+k\right)^{[k]}}{k!\left(\frac{n+5}{2}+k\right)^{[2]}}$$

(3.4.23) $\mathcal{E}(r^6) =$

$$\frac{15}{8}\sum_{0}^{\infty}\frac{\delta^{2k}}{(1+\delta^2)^{\left(\frac{n-1}{2}+k\right)}}\quad\frac{\left(\frac{n-3}{2}+k\right)^{[k]}}{k!\left(\frac{n+3}{2}+k\right)^{[3]}}$$

$$+\frac{45}{8}\mathcal{E}(\eta^2)\sum_{0}^{\infty}\frac{\delta^{2k+2}}{(1+\delta^2)^{\left(\frac{n+1}{2}+k\right)}}\quad\frac{\left(\frac{n-1}{2}+k\right)^{[k]}}{k!\left(\frac{n+5}{2}+k\right)^{[3]}}$$

$$+\frac{15}{4}\mathcal{E}(\eta^4)\sum_{0}^{\infty}\frac{\delta^{2k+4}}{(1+\delta^2)^{\left(\frac{n+3}{2}+k\right)}}\quad\frac{\left(\frac{n+1}{2}+k\right)^{[k]}}{k!\left(\frac{n+7}{2}+k\right)^{[3]}}$$

$$+\frac{1}{8}\mathcal{E}(\eta^6)\sum_{0}^{\infty}\frac{\delta^{2k+6}}{(1+\delta^2)^{\left(\frac{n+5}{2}+k\right)}}\quad\frac{\left(\frac{n+3}{2}+k\right)^{[k]}}{k!\left(\frac{n+9}{2}+k\right)^{[3]}}$$

(3.4.24) $\mathcal{E}(r^8) =$

$$\frac{105}{16}\sum_{0}^{\infty}\frac{\delta^{2k}}{(1+\delta^2)^{\left(\frac{n-1}{2}+k\right)}}\quad\frac{\left(\frac{n-3}{2}+k\right)^{[k]}}{k!\left(\frac{n+5}{2}+k\right)^{[4]}}$$

$$+\frac{105}{4}\mathcal{E}(\eta^2)\sum_{0}^{\infty}\frac{\delta^{2k+2}}{(1+\delta^2)^{\left(\frac{n+1}{2}+k\right)}}\quad\frac{\left(\frac{n-1}{2}+k\right)^{[k]}}{k!\left(\frac{n+7}{2}+k\right)^{[4]}}$$

$$+\frac{105}{8}\zeta(n^4)\sum_{0}^{\infty}\frac{\delta^{2k+4}}{(1+\delta^2)^{\left(\frac{n+3}{2}+k\right)}}\;\frac{\left(\frac{n+1}{2}+k\right)^{[k]}}{k!\left(\frac{n+9}{2}+k\right)^{[4]}}$$

$$+\frac{7}{4}\zeta(n^6)\sum_{0}^{\infty}\frac{\delta^{2k+6}}{(1+\delta^2)^{\left(\frac{n+5}{2}+k\right)}}\;\frac{\left(\frac{n+3}{2}+k\right)^{[k]}}{k!\left(\frac{n+11}{2}\right)^{[4]}}$$

$$+\frac{1}{16}\zeta(n^8)\;{}_{0}\;\frac{\delta^{2k+8}}{(1+\delta^2)^{\left(\frac{n+7}{2}+k\right)}}\;\frac{\left(\frac{n+5}{2}+k\right)^{[k]}}{k!\left(\frac{n+13}{2}+k\right)^{[4]}}$$

3.5 Moments of n^2. Case $k_1=n-1$.

Clearly moments of r^2 and thence R^2 are available for given $\delta^2=R_0^2/1-R_0^2$ when moments of η^2 are established; this section is devoted to establishing moments of η^2 for the case of two groups when one group contains n-1 individuals, the other group only one individual. Such a situation arises when only the "best" individual is picked out or singled out; the remaining (n-1) being lumped together in the category of "not best".

For convenience we replace $x_{01}<x_{02}<..<x_{0n}$ by $u_1<u_2<...<u_n$ so that we can consider $u_1<u_2<..<u_n$ to be a ranked set of n independently drawn observation from the standard normal population. With

$$(3.5.1) \qquad \bar{u} = \frac{1}{n} \sum_{1}^{n} u_i$$

then for $k_1 = n-1$

$$(3.5.2) \qquad \eta^2 = \frac{n}{n-1} (u_n - \bar{u})^2$$

The process of giving moments of η^2 in terms of moments of u_n is routine, however of some use is the fact that $u_n - \bar{u}$ and $\bar{u}$ have independent distributions.

We find after a little algebra:

$$(3.5.3) \qquad \mathcal{E}(\eta^2) = \frac{n}{n-1} \left\{ \mathcal{E}(u_n^2) - \frac{1}{n} \right\}$$

$$(3.5.4) \qquad \mathcal{E}(\eta^4) = \left(\frac{n}{n-1}\right)^2 \left\{ \mathcal{E}(u_n^4) - \frac{6}{n} \mathcal{E}(u_n^2) + \frac{3}{n^2} \right\}$$

$$(3.5.5) \qquad \mathcal{E}(\eta^6) = \left(\frac{n}{n-1}\right)^3 \left\{ \mathcal{E}(u_n^6) - \frac{15}{n} \mathcal{E}(u_n^4) + \frac{45}{n^2} \mathcal{E}(u_n^2) - \frac{15}{n^3} \right\}$$

$$(3.5.6) \qquad \mathcal{E}(\eta^8) = \left(\frac{n}{n-1}\right)^4 \left\{ \mathcal{E}(u_n^8) - \frac{28}{n} \mathcal{E}(u_n^6) + \frac{210}{n^2} \mathcal{E}(u_n^4) - \frac{420}{n^3} \mathcal{E}(u_n^2) + \frac{105}{n^4} \right\}$$

Since the first ten moments of the extreme (u_n) from a sample of n from a standard normal population are tabulated by Ruben for n=1(1)50 we give the first four moments of $1-r^2$ using equations (3.5.3) thru (3.5.6) and (3.4.21) thru (3.4.24) for the cases

$$\begin{cases} k_1=n-1; \quad k_2=1 \\ n=4(1)20(5)50 \\ R_o=0.05,\ 0.10(0.10)0.90, 0.95 \end{cases}$$

these moments which are of interest are given in Appendix A.

3.6 Moments of η^2; general two group situation.

In the general case of two groups we shall, for convenience put

$$k_1=n-k_2=k \tag{3.6.1}$$

It proved more convenient to work with a function.

$$t= \frac{nk^2}{(n+2)^3}\left(\frac{u_1+\dots+u_k}{k} - \bar{u}\right) \tag{3.6.2}$$

so that

$$\eta^2= \frac{(n+2)^3}{k(n-k)}\, t \tag{3.6.3}$$

Write

$$(3.6.4)\quad \bar{u}_{(k)} = \frac{1}{k}\sum_{1}^{k} u_j$$

then, since $\bar{u}_{(k)}-\bar{u}$ and $\bar{u}$ are independently distributed we find

$$(3.6.5)\quad \mathcal{E}(t) = \frac{nk^2}{(n+2)^3}\left\{\mathcal{E}(\bar{u}^2_{(k)}) - \frac{1}{n}\right\}.$$

$$(3.6.6)\quad \mathcal{E}(t^2) = \frac{n^2k^4}{(n+2)^6}\left\{\mathcal{E}(\bar{u}^4_{(k)}) - \frac{6}{n}\mathcal{E}(\bar{u}^2_{(k)}) + \frac{3}{n^2}\right\}$$

Using a method developed by Saw(1958) the moments of $\mathcal{E}(\bar{u}^h_{(k)})$ can be obtained as a power series in 1/(n+2). In fact if

$$(3.6.7)\quad \psi(p_k,n,a,b) = \psi_{a,}^{b} =$$

$$\mathcal{E}_{u_r}\left[\frac{e^{-\frac{1}{2}u_r^2}}{\int_{-\infty}^{u_r} e^{-\frac{1}{2}u^2}\,du}\right]^a u_r^b$$

when u_r has the distribution of the r^{th} largest of n standard normal variates and where

$$(3.6.8)\quad p_k = k/(n+1).$$

$\psi_{a,b}$ is used for $\psi(p_k,n,a,b)$ when n and k are fixed

then it is easily shown (following Saw, 1958) that

(3.6.9) $\mathcal{E}(t)=$

$$\frac{1}{(n+2)^3}\left(n\psi_{0,2}+n(k-1)(1-3\psi_{1,1})+n(k-1)(k-2)\psi_{2,0}-k^2 \right)$$

(3.6.10) $\mathcal{E}(t^2)=$

$$\frac{1}{(n+2)^6}\left(n^2\psi_{0,4}+n^2(k-1)(346\psi_{0,2}-11\psi_{1,1}-15\psi_{1,3}) \right.$$

$$+n^2(k-1)(k-2)(8\psi_{2,0}-18\psi_{1,1}+25\psi_{2,2}+3)$$

$$+n^2(k-1)(k-2)(k-3)(6\psi_{2,0}-10\psi_{3,1})$$

$$+n^2(k-1)(k-2)(k-3)(k-4)\psi_{4,0}$$

$$-6nk^2\psi_{0,2}-6nk^2(k-1)(1-3\psi_{1,1})$$

$$\left. -6nk^2(k-1)(k-2)\psi_{2,0}+k^4 \right)$$

For fixed values of $p_k=k/(n+1)$ we may write

$$(3.6.11) \quad \psi_{a,b}=\sum_{i=1}^{\ell} \frac{H_i(p_k,a,b)}{(n+2)^i} + 0(n+2)^{-(i+1)}$$

Values of $H_i(p_k,a,b)$ were originally given by Saw for $a+b\leq 4$, $i=0(1)5$ and $p_k=0.50(0.05)0.80$; however since those computation were made on a mechanical hand calculator the values were recalculated under this

project and the range extended to $a+b\leq 4$; $p_k=0.50(0.05)0.95$. The recomputed tables (which will be of use elsewhere in the theory of order statistics) are given in Appendix B.

Using (3.6.9), (3.6.10) the first two moments of t can be expressed as a power series in $\frac{1}{n+2}$ for a fixed value of p_k in fact if we may write

$$(3.6.12)\quad \mathcal{E}(t^s)=\sum_{j=0}^{\infty} J_j(p_k,s)(n+2)^{-j}$$

and values of $J_j(p_k,S)$ are given in Table I below for

$$\begin{cases} s=1,2 \\ p_k=0.50(0.05)0.95 \\ j=0(1)5 \end{cases}$$

(the choice $p_k\geq\frac{1}{2}$ is not restrictive since one group or the other must have at least half the observations contained in it).

The net result now is that for the case of two groups we have four moments of R^2 for general p when $k_1=n-1=n-k_2$ and two moments of R^2 for general $\bar{p}$ and general $\bar{k}_1,(1<k_1<n-1)$. On the basis of these moments we give in the next chapter an indication of the power of our proposed method and find that the power is suprisingly good.

Table I

Values of $J_j(\frac{p_k}{2}, s=1)$. (see equation 3.6.12)

p_k \ j	0	1	2	3	4	5
.50	.159154942259	-1.11408459495	2.4928286094	-1.650592081	-.07347003	-.1647128
.55	.15661499829	-1.15240713187	2.6593750631	-1.752995266	-.04898468	-.1748879
.60	.149260553713	-1.16644960245	2.8014808876	-1.37333157	-.00795731	-.1625964
.65	.137195421889	-1.15494631228	2.9192947846	-1.896905892	.06751073	-.0970113
.70	.120890154750	-1.11667628024	3.0145159074	-1.919950028	.21381524	-.0952681
.75	.100981941712	-1.05024714838	3.0915026182	-1.883492631	.51351459	.6076852
.80	.078378676188	- .95368669545	3.1499063813	-1.736893876	1.18083092	2.0319833
.85	.054362995649	- .82354965882	3.2418224878	-1.348830659	2.8919384	6.6491090
.90	.030799654052	- .65255023373	3.3961132584	- .278872213	8.60567987	27.5692897
.95	.010636959720	- .42094588636	3.8454744689	2.992368438	47.01506832	374.0766422

Table I

Values of J_j (p_k, s=2) (see equation 3.6.12)

p_k \ j	0	1	2	3	4	5
.50	.025330296019	-.29679038748	1.3913320308	-3.219769636	3.65138528	-1.9094930
.55	.024542825789	-.30481978970	1.5005386292	-3.614757849	4.25137133	-2.3222867
.60	.022278712935	-.29652511480	1.5487083569	-3.913649434	4.78200624	-2.7296553
.65	.018822583773	-.27230982249	1.5291460906	-4.095082518	5.21501971	-3.1197959
.70	.014514429557	-.23424219463	1.4398479278	-4.1442212[illegible]9	5.52603165	-3.4844550
.75	.010197354264	-.18599803449	1.2837318713	-4.053598352	5.69573239	-3.8192926
.80	.006143216900	-.13269915946	.10686393559	-3.823444081	5.71218658	-4.1182426
.85	.002955335325	-.08063891609	.8071940073	-3.460763846	5.57983943	-4.3288041
.90	.020948618699	-.03685969059	.5169222685	-2.973203782	5.37064823	-3.9995064
.95	.000113144913	-.00841491226	.2230299519	-2.326794502	5.41201679	18.95385873

CHAPTER 4.

THE POWER OF THE TEST OF $R_o^2=0$.

4.1 Introduction

We consider the case of two groups only (i.e. c=2). It is to be expected that, all else being equal, as the number of groups increases, the power of the test will also increase since essentially more information is at the statisticians disposal. When c=n we reach the case discussed in part I of this contract when it was found that the efficiency of the test based on a complete set of rankings was very close to unity. This was the case when $d_j=(j-\frac{n+1}{2})/(\frac{n(n^2-1)}{12})^{1/2}$ $j=1,2,\ldots,n$.

4.2 Large sample behaviour

Taking the case p=1; c=2 and general n we consider the conditional moment generating function of $-2\rho\log(1-r^2)$ where ρ is some arbitrary constant and r is the correlation between the d_j and the (scalar) normally distributed observations x_{ij}. We write

$$(4.2.1)\quad \phi^{(\theta)}_{-2\rho\log(1-r^2)|n^2/s^2} = \mathcal{E}\{\exp(-2\rho\theta\log(1-r^2))|n^2/s^2\}$$

$$= \mathcal{E}\{(1-r^2)^{-2\rho\theta}|n^2/s^2\}.$$

Required then is the conditional $(-2\rho\theta)$-th moment of $1-r^2$ which is available using the work of section 3.4. In fact

$$(4.2.2)\quad \phi_{-2\rho\log(1-r^2)\,|\,\eta^2/s^2}(\theta) = (1-r_o^2)^{\frac{n-1}{2}} \sum_{ij}\sum r_o^{2i+2j} \frac{\left(\frac{\eta^2}{s^2}\right)^i \left(1-\eta^2/s^2\right)^j}{i!j!} \frac{\Gamma\left(\frac{n-1}{2}+i+j\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{B\left(i+\frac{1}{2};\ \frac{v}{2}+j-2\rho\theta\right)}{B\left(i+\frac{1}{2};\ \frac{v}{2}+j\right)}$$

where $v=n-2$.

Now, using the expansion for $\log \Gamma(x+h)$ in terms of the Bernoulli polynomials in h (see for example Part 4, of volume I of this contract, in particular page 94) we have

$$(4.2.3)\quad \log \frac{B\left(i+\frac{1}{2};\ \frac{v}{2}+j-2\rho\theta\right)}{B\left(i+\frac{1}{2};\ \frac{v}{2}+j\right)} = -\left(i+\tfrac{1}{2}\right)\log(1-2\theta) + \sum_{r=1}^{\infty} \frac{\pi_r}{\rho^r}\left\{\frac{1}{(1-2\theta)^r} - 1\right\}$$

with

$$(4.2.4)\quad \pi_r = \frac{(-)^r}{r(r+1)}\left\{B_{r+1}\left(\frac{v+1}{2} + i+j-\rho\right) - B_{r+1}\left(\frac{v}{2}+j-\rho\right)\right\}$$

and $B_s(h)$ the sth Bernoulli polynomial in h.

In particular.

$$(4.2.5)\quad \pi_1 = -\tfrac{1}{2}\left(i+\tfrac{1}{2}\right)\left(v+2j+i-2\rho-\tfrac{1}{2}\right)$$

$$(4.2.6) \qquad \pi_2 = \frac{1}{6}(i+\tfrac{1}{2})\left\{ (v-\tfrac{1}{2}+i+2j-2\rho)^2 + \tfrac{1}{2}(v-\tfrac{1}{2}+i+2j-2\rho) - (\tfrac{v+1}{2}+i+j-\rho)(\tfrac{v}{2}+j-\rho)+\tfrac{1}{2} \right\}$$

If we now set

$$(4.2.7) \qquad \rho = \frac{1}{2}v + a$$

with a of order zero in n then

$$(4.2.8) \qquad \pi_r = \frac{(-)^r}{r(r+1)} \left\{ B_{r+1}(\tfrac{1}{2}+i+j-a) - B_{r+1}(j-a) \right\}$$

so that π_r is of order zero in n.

With ρ defined by (4.2.7) we get

$$(4.2.9) \qquad \frac{B(i+\frac{1}{2};\ \frac{v}{2}+j-2\rho\theta)}{B(i+\frac{1}{2},\ \frac{v}{2}+j)} = \frac{1}{(1-2\theta)^{i+\frac{1}{2}}} \left[1 + \frac{2\theta}{(1-2\theta)} \left(\frac{\pi_1}{\rho} \right) + O(\frac{1}{\rho^2}) \right]$$

and using this in (4.22) after performing the summations over i and j

$$(4.2.10) \qquad \phi_{-2\rho \log (1-r^2)}(\theta) \Big|_{n^2/s^2} =$$

$$\frac{(1-2\theta)^{\frac{n-2}{2}}}{\left\{(1-2\theta)-2\theta\frac{\delta^2 n^2}{s^2}\right\}^{\frac{n-1}{2}}} + \frac{1}{\rho}(n\,r_o^2)^2 b_1$$

$$+ \frac{1}{\rho}(nr_o^2)b_2$$

$$+ \frac{1}{\rho} b_3 \quad +O(\frac{1}{\rho^2})$$

where b_1 b_2 and b_3 converge to limits which are constant in θ and $\delta^2 n^2/s^2$ as n approaches infinity. For large n therefore since ρ behaves as n, the first term of (4.2.10) describes the behaviour of the moment generating function of $-2\rho\log(1-r^2)$ conditional on n^2/s^2.

Allowing n to go to infinity and $\delta^2=r_o^2/(1-r_o^2)$ to go to zero in such a manner that

(4.2.11) $\lim\limits_{n\to\infty} n\delta^2 = \lambda.$

we have

(4.2.12) $\lim\limits_{\substack{n\to\infty \\ n\delta^2=\lambda}} \phi_{-2\rho\log(1-r^2)\,|\,n^2/s^2}(\theta) =$

$$(1-2\theta)^{-\frac{1}{2}}\exp\left\{\frac{\theta}{(1-2\theta)}\frac{\lambda n^2}{s^2}\right\}$$

Conditional on η^2/s^2 therefore, $-2\rho\log(1-r^2)$ is asymptotically distributed as the square of a normal variate with mean

$$\left(\frac{\lambda\eta^2}{s^2}\right)^{1/2}$$ and unit variance.

Thus $\Pr\{r^2>\alpha \mid \eta^2/s^2\}$ is an increasing function of λ for any η^2/s^2 and thus $\Pr\{r^2>\alpha\}$ is an increasing function λ. (non-conditionally on η^2/s^2).

4.3 Small sample behaviour

To discuss the small sample behaviour, it is of interest to consider the regular product - moment correlation coefficient between the set of variables $\{(u_i,v_i)\}_{i=1}^{i=n}$ where.

$$(4.3.1)\quad \begin{pmatrix} u_i \\ v_i \end{pmatrix} \frown N_2\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix} ; \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right\} \qquad i=1,2,\ldots,n$$

The product moment estimate of ρ, call this $\hat{\rho}$, is given as

$$(4.3.2)\quad \hat{\rho} = \frac{\sum_i (u_i-\bar{u})^2 (v_i-\bar{v})^2}{\left\{\sum_i (u_i-\bar{u})^2 \sum_i (v_j-\bar{v})^2\right\}^{1/2}}$$

Now set

$$(4.3.3)\quad \alpha_i = \frac{u_i-\bar{u}}{\{\Sigma(u_i-\bar{u})^2\}^{1/2}} \qquad i=1,2,\ldots,n$$

so that

$$(4.3.4)\quad \hat{\rho}^2 = \frac{\left[\Sigma\alpha_i (v_i-\bar{v})\right]^2}{\left[\Sigma(v_i-\bar{v})^2\right]} = p^2/(p^2+q^2).$$

where

(4.3.5) $\hat{p} = \Sigma\alpha_i(v_i-\bar{v})$

(4.3.6) $q^2=\Sigma(v_i-\bar{v})^2-\left[\Sigma\alpha_i(v_i-\bar{v})\right]^2$

Now

(4.3.7) $v_i|u_i \frown N_1(\rho u_i;\ 1-\rho^2)$

so that

(4.3.8) $p|\{u_i\}_1^n \frown N_1\{\rho\Sigma\alpha_i(u_i-\bar{u})^2;\ 1-\rho^2\}$

or

(4.3.9) $p^2|\{u_i\}_1^n \frown (1-\rho^2)\chi_1^2(\delta^2 s^2)$

where

(4.3.10) $\delta^2=\rho^2/(1-\rho^2)$

and

(4.3.11) $s^2=\sum_{i=1}^{n}(u_i-\bar{u})^2$

Also

(4.3.12) $q^2|\{u_i\}_1^n \quad (1-\rho^2)\chi_{n-2}^2(0)$

and p^2 and q^2 are independently distributed condition on $\{u_i\}_1^n$. (this by the standard theory of least squares).

The distribution of $\hat{\rho}^2=p^2/(p^2+q^2)$ is the same as that for r^2 except that for the case of $\hat{\rho}^2$, η^2 becomes identically equal to s^2. Conditional on $\{u_i\}_1^n$ therefore r^2 behaves as $\hat{\rho}^2$ except that δ^2 has to be replaced by $\delta^2\eta^2/_s2$. It follows from the properties of the regular correlation coefficient (or rather its square :-$\hat{\rho}^2$) that

$$(4.3.13)\quad \begin{cases} Pr\{r^2>\alpha\,|\,\eta^2/_{s}2\} & \text{increases with } r_o^2;\ n \text{ fixed} \\ Pr\{r^2>\alpha\,|\,\eta^2/_{s}2\} & \text{increases with } n;\ r_o^2 \text{ fixed.} \end{cases}$$

It has been easy to discuss this in the context when p=1. Obviously, by comparing our R^2 with the estimate $R^2_{o(1,2,..,p)}$ of the square of the regular multiple correlation coefficient between x_o and $(x_1, x_2,\ldots x_p)$ we draw the same conclusions for general dimension p.

Finally then if $\{d_j;\underline{x}_j\}_{j=1}^{j=n}$ is the observed data drawn from a population with multiple correlation R_o (that is: the multiple correlation between the unobserved x_{oj} and the $\underline{x}_j$) then if

$$\hat{R} = \max_{(\{d_j\};\underline{\alpha})} \text{corr}\ (d_j,\ \underline{\beta}'\underline{x}_j)$$

as defined throughout this work, then

$$Pr\{\hat{R}^2>\alpha\}\quad \begin{matrix} \text{increases with } R_o^2,\ n \text{ fixed} \\ \text{increases with } n\ ,\ R_o \text{ fixed} \end{matrix}$$

4.4 Numerical values of power in a specific case.

Since we have moments of R^2 under the alternative hypothesis it is possible to fit a frequency curve to the

moments of R^2 to obtain an approximation to the power. It turns out that a Type I Pearson curve provides a good fit as was to be anticipated. The case n=15 was considered:-

Power of the size 0.05 test of H_o $R_o=0$

n=15 c=2 $k_1+k_2=15$

R_o	$k_1=14$.*	$k_1=12$ ($p_k=0.75$).**	$k_1=8$.** ($p_k=0.50$)
0.00	0.05	0.05	0.05
0.05	0.05	0.05	0.05
0.10	0.05	0.06	0.06
0.20	0.06	0.07	0.08
0.30	0.07	0.10	0.15
0.40	0.10	0.16	0.23
0.50	0.14	0.21	0.33
0.60	0.18	0.28	0.44
0.70	0.21	0.41	0.57
0.80	0.26	0.66	0.81
0.90	0.31	0.83	0.95
0.95	0.37	0.93	0.99

(*four moment fit; see section 3.5)

(**two moment fit; see section 3.6)

The power, at least for n=15, is not high for the case of two groups when one group has only one member in it; considering the almost complete lack of information (on R_o^2) which this situation represents, it is perhaps surprising that R^2 is of any use at all. For a more reasonable dichotomy of individuals ($p_k=0.75$ and $p_r=0.50$) the power is quite high. It is noted

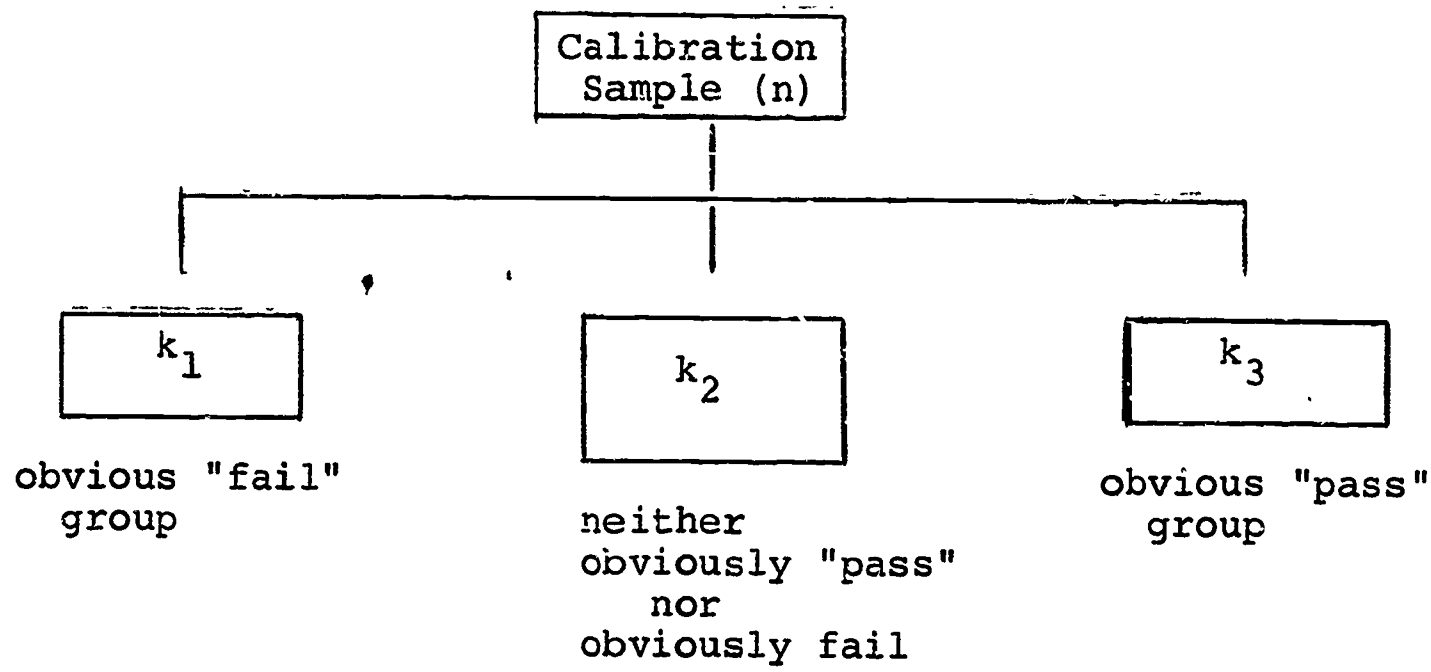
that that for given R_o^2 the power in the case p_k=0.50 is greater than the corresponding figure for p_k=0.75 which in turn is higher than the corresponding figure for p_r=7/8 (case k_1=14). This confirms what one would anticipate; that for the case of two groups, the most information is to be had when the group sizes are equal.

SUMMARY

We have presented a possible approach to the problem of classification of an individual on the basis of a vector of scores when the usual classical model does not hold. It is our contention that within a group homogeneous with regard to background <u>there</u> <u>does</u> <u>not</u> <u>exist</u> two distinct populations:- a "fail" population and a "pass" population; rather there is a continuous spectrum: the upper end of which are more likely to pass than those members to be found at the lower end of the spectrum.

Making use of this new model we are able to give a method of evaluating the vector of scores (on admission tests possibly) in respect of their value as indicators of the ultimate worth or performance of the individual. The efficiency of the method is surprisingly high.

It is noted that in addition to the critisms of the use of the classical discriminent approach leveled above; the classical approach requires a large calibration sample (that is in our notation n must be large) in order that the classical theory may hold approximately for in the classical theory we must first pick out those individuals which cannot be readily assigned to a group. In our model, the individuals who cannot be readily assigned form a (middle) group and the data relevant to such an individual is likely to be of great value. Diagramitically:-



Classical model:- makes use of k_1+k_3 observations.

Our model:- makes use of $k_1+k_2+k_3$ observations.

Experience in the field yields the fact that frequently k_2 is larger even than k_1+k_3.

Interestingly enough, if it were possible to use all n members in the classical theory then the test function and discriminant vector is algebraically the same though the distribution of these quantities is different under the two models.

Given the discriminant vector ($\hat{\underline{\beta}}$) under our model, it may be treated as in the classical case to solve the problems.

(a) Compare a new group of individuals with respect to their predicted future performance.

(b) evaluate the choice of admission tests (for example) as predictors of future performances.

APPENDIX A

Moments of $1-r^2$

(see Section 3.4; in particular equations

(3.4.21) thru (3.4.24).

Cases:-

Sample size $n=4(1)20(5)50$

Two groups $c=2.$

Group sizes $k_1=n-1$

$k_2=1$

Vector dimension $p=1$ *

Population correlation $R_o=0.05,\ 0.10(0.10)\ 0.90,\ 0.95$

(*The moments for general p are obtainable from the results for p=1 using (3.3.12)).

$R_o = .05$

N	μ_1'	μ_2	μ_3	μ_4
4	.66629885	.08895878	-.01690310	.01693457
5	.74958306	.06260402	-.01562485	.01173170
6	.79956186	.04583089	-.01220669	.00799750
7	.83288820	.03484069	-.00928319	.00551737
8	.85669835	.02732611	-.00709449	.00388681
9	.87456033	.02198470	-.00549568	.00280049
10	.88845637	.01806065	-.00432333	.00206177
11	.89957588	.01509675	-.00345275	.00154825
12	.90867579	.01280494	-.00279640	.00118361
13	.91626079	.01099708	-.00229391	.00091951
14	.92268027	.00954626	-.00190356	.00072473
15	.92818386	.00836449	.00159617	.00057867
16	.93295460	.00738922	-.00135107	.00046749
17	.93712984	.00657506	-.00115340	.00038171
18	.94081456	.00588841	-.00099228	.00031468
19	.94409048	.00530400	-.00085971	.00026172
20	.94702208	.00480249	-.00074966	.00021942
25	.95802067	.00311244	-.00040932	.00010046
30	.96523199	.00218010	-.00024731	.00005226
35	.97032557	.00161189	-.00016065	.00002981
40	.97411523	.00124017	-.00011018	.00001822
45	.97704506	.00098375	-.00007883	.00001176
50	.97937802	.00079943	-.00005833	.00000792

$R_o = .10$

N	μ_1'	μ_2	μ_3	μ_4
4	.66519224	.08916696	-.01681743	.01694415
5	.74832911	.06291431	-.01562261	.01176969
6	.79824454	.04617874	-.01225336	.00805132
7	.83155015	.03519404	-.00935303	.00557540
8	.85536243	.02766973	-.00717279	.00394269
9	.87323911	.02231176	-.00557472	.00285149
10	.88715682	.01836858	-.00439926	.00210703
11	.89830171	.01538508	-.00352384	.00158786
12	.90742881	.01307427	-.00286201	.00121804
13	.91504164	.01124849	-.00235400	.00094937
14	.92148893	.00978103	-.00195838	.00075061
15	.92701985	.00858392	-.00164610	.00060114
16	.93181723	.00759459	-.00139653	.00048705
17	.93601824	.00676757	-.00119481	.00039878
18	.93972782	.00606915	-.00103006	.00032963
19	.94302763	.00547397	-.00089422	.00027484
20	.94598217	.00496260	-.00078124	.00023099
25	.95708204	.00323404	-.00043018	.00010692
30	.96437571	.00227569	-.00026174	.00005613
35	.96953717	.00168915	-.00017105	.00003227
40	.97338367	.00130403	-.00011792	.00001986
45	.97636181	.00103750	-.00008475	.00001289
50	.97873636	.00084537	-.00006297	.00000873

2

$R_o = .20$

N	μ_1'	μ_2	μ_3	μ_4
4	.66071740	.08997692	-.01645475	.01697529
5	.74326541	.06412813	-.01558670	.01190896
6	.79293059	.04754000	-.01241054	.00825149
7	.82615717	.03657602	-.00960289	.00579218
8	.84998176	.02901256	-.00745794	.00415186
9	.86792080	.02358888	-.00586483	.00304250
10	.88192827	.01957007	-.00467915	.00227664
11	.89317741	.01650935	-.00378651	.00173633
12	.90241566	.01412382	-.00310484	.00134712
13	.91014197	.01222767	-.00257666	.00106129
14	.91670233	.01069492	-.00216165	.00084765
15	.92234426	.00943775	-.00183133	.00068541
16	.92724962	.00839338	-.00156525	.00056039
17	.93155504	.00751604	-.00134856	.00046279
18	.93536518	.00677163	-.00117033	.00038568
19	.93876162	.00613440	-.00102238	.00032409
20	.94180885	.00558455	-.00089854	.00027440
25	.95331741	.00370587	-.00050769	.00013117
30	.96094278	.00264631	-.00031539	.00007063
35	.96637732	.00198853	-.00020968	.00004152
40	.97045230	.00155137	-.00014669	.00002603
45	.97362450	.00124564	-.00010676	.00001717
50	.97616605	.00102320	-.00008021	.00001180

$\underline{R_o = .30}$

N	μ_1'	μ_2	μ_3	μ_4
4	.65308021	.09124084	-.01577892	.01700190
5	.73464938	.06604872	-.01543113	.01209680
6	.78390998	.04969647	-.01256816	.00853274
7	.81701954	.038762288	-.00991561	.00610067
8	.84087892	.03113398	-.00783478	.00445112
9	.85893486	.02560307	-.00625729	.00331649
10	.87310345	.02146193	-.00506251	.00252028
11	.88453648	.01827702	-.00414898	.00194928
12	.89396885	.01577178	-.00344155	.00153278
13	.90189209	.01376330	-.00288639	.00122235
14	.90864776	.01212663	-.00244509	.00098738
15	.91448075	.01077405	-.00209004	.00080676
16	.91957142	.00964246	-.00180120	.00066605
17	.92405561	.00868549	-.00156380	.00055506
18	.92803760	.00786844	-.00136684	.00046650
19	.93159889	.00716487	-.00120205	.00039512
20	.93480405	.00655440	-.00106306	.00033705
25	.94700660	.00443989	-.00061658	.00016624
30	.95519311	.00322193	-.00039081	.00009178
35	.96108847	.00245295	-.00026404	.00005498
40	.96554832	.00193472	-.00018718	.00003502
45	.96904697	.00156801	-.00013776	.00002342
50	.97186917	.00129848	-.00010449	.00001628

$R_o = .40$

N	μ_1'	μ_2	μ_3	μ_4
4	.64198408	.09280952	-.01468235	.01698691
5	.72219168	.06850159	-.01501310	.01227014
6	.77091606	.05245990	-.01257254	.00882163
7	.80389620	.04156191	-.01014011	.00642766
8	.82783712	.03384293	-.00816136	.00477265
9	.84608633	.02816856	-.00662182	.00361295
10	.86050658	.02386575	-.00543127	.00278500
11	.87221984	.02051797	-.00450490	.00218234
12	.88194385	.01785672	-.00377661	.00173548
13	.89016014	.01570252	-.00319745	.00139847
14	.89720443	.01393161	-.00273163	.00114022
15	.90331829	.01245623	-.00235289	.00093979
16	.90868022	.01121269	-.00204184	.00078200
17	.91342516	.01015382	-.00178398	.00065641
18	.91765707	.00924401	-.00156838	.00055537
19	.92145749	.00845594	-.00138668	.00047331
20	.92489124	.00776837	-.00123242	.00040606
25	.9380934	.00535533	-.00072931	.00020507
30	.94708352	.00393801	-.00046915	.00011524
35	.95363626	.00302963	-.00032062	.00007000
40	.95864367	.00241008	-.00022939	.00004510
45	.96260579	.00196731	-.00017012	.00003045
50	.96582584	.00163916	-.00012986	.00002134

$R_o = .50$

N	μ_1'	μ_2	μ_3	μ_4
4	.62694719	.09442490	-.01301595	.01687940
5	.70542953	.07119261	-.01413762	.01234705
6	.75352781	.05552012	-.01221773	.00902538
7	.78641044	.04466144	-.01007743	.00668244
8	.81052067	.03683511	-.00825350	.00503366
9	.82907557	.03099326	-.00679155	.00385888
10	.84386925	.02650391	-.00563587	.00300751
11	.85598596	.02296990	-.00472085	.00237956
12	.86612229	.02013155	-.00399115	.00190851
13	.87474782	.01781300	-.00340388	.00154957
14	.88219147	.01589146	-.00292668	.00127211
15	.88869117	.01427893	-.00253524	.00105481
16	.89442363	.01291091	-.00221124	.00088257
17	.89952304	.01173912	-.00194079	.00074457
18	.90409336	.01072686	-.00171327	.00063288
19	.90821644	.00984572	-.0015204	.00054166
20	.91195771	.00907347	-.00135595	.00046653
25	.92649585	.00633451	-.00081335	.00023942
30	.93655130	.00470128	-.00052829	.00013617
35	.94397103	.00364276	-.00036369	.00008349
40	.94969792	.00291450	-.00026171	.00005422
45	.95426738	.00239039	-.00019501	.00003684
50	.95800762	.00199968	-.00014945	.00002596

$R_o = .60$

N	μ_1'	μ_2	μ_3	μ_4
4	.60720292	.09564254	-.01060029	.01661229
5	.68364340	.07363308	-.01257337	.01223138
6	.73110269	.05837772	-.01126605	.00904058
7	.76399639	.04757244	-.00950388	.00676714
8	.78843174	.03964300	-.00790885	.00514691
9	.80746322	.03363618	-.00658680	.00397867
10	.82280163	.02896346	-.00551798	.00312315
11	.83548700	.02524751	-.00465769	.00248641
12	.84619203	.02223733	-.00396284	.00200503
13	.85537339	.01976033	-.00339791	.00163572
14	.86335334	.01769447	-.00293498	.00134856
15	.87036653	.01595127	-.00255257	.00112247
16	.87658847	.01446520	-.00223415	.00094243
17	.88215335	.01318683	-.00196698	.00079757
18	.88716572	.01207825	-.00174119	.00067988
19	.89170844	.01110995	-.00154908	.00058344
20	.89584812	.01025865	-.00138459	.00050375
25	.91210216	.00721778	-.00083781	.00026115
30	.92351184	.00538659	-.00054748	.00014966
35	.93202620	.00419140	-.00037859	.00009233
40	.93865716	.00336470	-.00027338	.00006026
45	.94398704	.00276720	-.00020428	.00004112
50	.94837681	.00232024	-.00015692	.00002909

$R_o = .70$

N	μ_1'	μ_2	μ_3	μ_4
4	.58149081	.09566119	-.00726825	.01608673
5	.65568920	.07498365	-.01010240	.01180990
6	.70264670	.06021020	-.00950174	.00875734
7	.73579809	.04951680	-.00822473	.00658402
8	.76083152	.04154363	-.00695800	.00502883
9	.78060750	.03543138	-.00586227	.00390200
10	.79674203	.03063316	-.00495347	.00307303
11	.81022763	.02679018	-.00420921	.00245352
12	.82171309	.02365937	-.00360047	.00198350
13	.83164349	.02107118	-.00310075	.00162179
14	.84033606	.01890432	-.00268815	.00133975
15	.84802405	.01707000	-.00234520	.00111716
16	.85488341	.01550197	-.00205816	.00093952
17	.86104983	.01414991	-.00181628	.00079630
18	.86662981	.01297500	-.00161110	.00067975
19	.87170837	.01194691	-.00143597	.00058407
20	.87635428	.01104159	-.00128558	.00050491
25	.89476517	.00779618	-.00078267	.00026305
30	.90785494	.00583254	-.00051349	.00015133
35	.91771584	.00454670	-.00035609	.00009364
40	.92545222	.00365513	-.00025767	.000061[illegible]7
45	.93170779	.00300953	-.00019285	.00004190
50	.93688553	.00252583	-.00014832	.00002970

$R_o = .80$

N	μ_1'	μ_2	μ_3	μ_4
4	.54755326	.09289019	-.00302091	.01509842
5	.61961834	.07368696	-.00667500	.01091016
6	.66653473	.05957676	-.00687570	.00803555
7	.70046233	.04919442	-.00620323	.00602180
8	.72658416	.04137517	-.00537659	.00459141
9	.74754462	.03534283	-.00460324	.00355895
10	.76486361	.03058774	-.00393447	.00280104
11	.77949102	.02676872	-.00337221	.00223540
12	.79205805	.02365145	-.00290390	.00180661
13	.80300418	.02107099	-.00251428	.00147682
14	.81264682	.01890847	-.00218927	.00121978
15	.82122222	.01707651	-.00191690	.00101698
16	.82891045	.01550965	-.00168743	.00085517
17	.83585167	.01415804	-.00149299	.00072474
18	.84215688	.01298316	-.00132730	.00061861
19	.84791525	.01195483	-.001185330	.00053150
20	.85319948	.01104913	-.00106296	.00045942
25	.87429365	.00780126	-.00065095	.00023928
30	.88943968	.00583553	-.00042851	.00013761
35	.90093117	.00454821	-.00029775	.00008513
40	.90999634	.00365561	-.00021572	.00005569
45	.91735857	.00300930	-.00016157	.00003807
50	.92347440	.00252511	-.00012432	.00002697

$R_o = .90$

N	μ_1'	μ_2	μ_3	μ_4
4	.50054909	.08351268	.00131107	.01295930
5	.57159125	.06638048	-.00295492	.00906511
6	.61977443	.05362577	-.00393690	.00655672
7	.65563070	.04420761	-.00389605	.00486030
8	.68379691	.03711500	-.00354542	.00367949
9	.70672710	.03165066	-.00312845	.00283785
10	.72587923	.02735075	-.00272924	.00222519
11	.74219015	.02390334	-.00237387	.00177064
12	.75629644	.02109398	-.00206682	.00142758
13	.76864955	.01877184	-.00180474	.00116462
14	.77958040	.01682838	-.00158197	.00096022
15	.78933834	.01518396	-.00139258	.00079931
16	.79811536	.01377902	-.00123120	.00067117
17	.80606207	.01256826	-.00109322	.00056805
18	.81329870	.01151674	-.00097476	.00048427
19	.81992249	.01059713	-.00087263	.00041559
20	.82601307	.00978779	-.00078417	.00035884
25	.85043922	.00689070	-.00048329	.00018598
30	.86808662	.00514197	-.00031899	.00010651
35	.88153582	.00399923	-.00022184	.00006564
40	.89218191	.00320836	-.00016071	.00004279
45	.90085202	.00263665	-.00012029	.00002916
50	.90807089	.00220898	-.00009247	.00002060

$R_o = .95$

N	μ_1'	μ_2	μ_3	μ_4
4	.46833526	.07304560	.00224984	.01061373
5	.54042725	.05795840	-.00189192	.00729577
6	.59052062	.04675914	-.00301085	.00524249
7	.62828533	.03851306	-.00312577	.00387866
8	.65817054	.03231503	-.00290829	.00293618
9	.68261051	.02754534	-.00259868	.00226599
10	.70308464	.02379441	-.00228504	.00177823
11	.72055807	.02078815	-.00199808	.00141611
12	.73569342	.01833871	-.00174610	.00114252
13	.74896378	.01631423	-.00152878	.00093258
14	.76071787	.01461999	-.00134270	.00076923
15	.77121937	.01318651	-.00118369	.00064051
16	.78067180	.01196184	-.00104767	.00053792
17	.78923533	.01090651	-.00093101	.00045531
18	.79703792	.00999005	-.00083063	.00038815
19	.80418322	.00918865	-.00074391	.00033307
20	.81075625	.00848343	-.00066871	.00028754
25	.83714535	.00596016	-.00041224	.00014881
30	.85623925	.00443864	-.00027188	.00008504
35	.87080740	.00344556	-.00018885	.00005228
40	.88234989	.00275915	-.00013662	.00003399
45	.89175730	.00226358	-.00010210	.00002311
50	.89959528	.00189334	-.00007837	.00001628

APPENDIX B

Values of

$$\psi(Pr;\ a,b) = \left[\frac{Z(x_{(r)})}{P(x_{(r)})}\right]^{a} x_{(r)}^{\ b}$$

where $x_{(1)}<x_{(2)}<\ldots<x_{(r)}<\ldots<x_{(n)}$ is an ordered sample from the standard normal density with p.d.f. $Z(x)$ and c.d.f. $P(x)$. (See section 3.6).

We write

$$(Pr;\ a,b) = \underset{i=0}{} H_i(Pr;\ a,b)\ (n+2)^{-i}$$

and tabulate

$H_i(Pr;\ a,b)$ $\qquad$ $i=0(1)5$

$Pr=0.50(0.05)0.95$

$1\leq a+b\leq 4.$

Values of $H_i(p_r: 1, 0)$

$p_r \backslash i$	0	1	2	3	4	5
.50	.797884560803	.17122749215	.2675948210	.351805280	.35083227	.2064217
.55	.719645233605	.12315255795	.2087024679	.295036278	.31199406	.1824069
.60	.643904222437	.08049221674	.1608872152	.253357686	.27780583	.1288795
.65	.569844622135	.04185619013	.1225515517	.226052169	.24594889	.0328819
.70	.496703734572	.00620055709	.0933757847	.214321023	.20896424	-.1434091
.75	.423702096902	-.02729418604	.0746159386	.222054807	.14575210	-.4957199
.80	.349952400454	-.05929802616	.0703206914	.257981643	-.00997784	-1.2871582
.85	.274304394122	-.09036257256	.0912810908	.341417162	-.51387773	-3.3818363
.90	.194998114361	-.12084195002	.1707336658	.519024357	-2.79336338	-10.5526206
.95	.108563828827	-.15011410867	.4697827656	.891805119	-25.41847592	-43.9458986

$H_i(p_r: 0, 1)$

p_r \ i	0	1	2	3	4	5
.50	zero	zero	zero	zero	zero	zero
.55	.125661346949	.09926236825	.1408964190	.155228947	.09799231	-.0221991
.60	.253347103490	.20368176169	.2915866851	.319735637	.18602288	-.0837663
.65	.385320466488	.31947278036	.4644317213	.504596336	.24415985	-.2477102
.70	.524400512708	.45547178183	.6784530714	.725083349	.22009660	-.6469362
.75	.674489750232	.62618531327	.9682126906	1.004304934	-.04298873	-1.6093112
.80	.841621233763	.85903081557	1.4077295811	1.377353483	-1.09851674	-4.0542181
.85	1.036433556055	1.21539731829	2.1918620707	1.876920385	-5.48141507	-10.8161668
.90	1.281551694520	1.87241797056	4.0294326977	2.226019824	-31.72733374	-25.7922117
.95	1.644853644578	3.67259773138	11.9815222869	-6.674482002	-482.31284872	825.4817069

$H_i(p_r: 2, 0)$

p_r \ i	0	1	2	3	4	5
.50	.636619772367	.90985931710	1.3346947489	1.779126085	2.06923999	1.9773401
.55	.517889262250	.76187856959	1.0699405655	1.390360023	1.64363754	1.6897329
.60	.414612647672	.64036518438	.8554832033	1.080421590	1.30928959	1.4428607
.65	.324722893376	.53896323628	.6765219822	.826910013	1.05226379	1.2449510
.70	.246714599937	.45301134421	.5222717057	.615532357	.86991402	1.0878890
.75	.179523466919	.37887923665	.3840326688	.438483664	.77579515	.9247723
.80	.122466682583	.31345895824	.2536049522	.296262624	.81501225	.5550833
.85	.075242900634	.25360876457	.1212633219	.208651333	1.11143498	-.9842278
.90	.038024264604	.19511646612	-.0285918745	.268690727	2.03425174	-10.0813625
.95	.011786104929	.12922043499	-.2330995168	.001279544	.04739377	.2264453

$H_i(p_r: 1, 1)$

p_r \ i	0	1	2	3	4	5
.50	zero	-1.00000000000	-1.4292036732	-1.820209919	-1.[illegible]789869	-1.8634579
.55	.090431589380	-.87413999205	-1.1834485522	-1.422807678	-1.52248776	-1.5658530
.60	.163131269680	-.77742574309	-.9874295998	-1.091268202	-1.11785951	-1.3548757
.65	.219572795627	-.70438333756	-.825191972412	-.794575255	-.76370379	-1.3106205
.70	.260471693073	-.65155438828	-.6834646609	-.502705751	-.45455268	-1.6102898
.75	.285782721512	-.61705901234	-.5480300816	-.176862435	-.22251886	-2.7275573
.80	.294527371028	-.60052637391	-.3977251125	.250199712	-.23533181	-6.1433253
.85	.284298278641	-.60351421565	-.1879863576	.931380993	-1.30149235	-17.6939877
.90	.249900163887	-.63109429048	.2151006010	2.341288774	-8.72695960	-71.8750015
.95	.178571609515	-.69825858915	1.5105138909	7.043482021	-103.94267227	-678.8559724

$H_i(p_r: 0, 2)$

p_r \ i	0	1	2	3	4	5
.50	zero	1.57059632679	2.4674011002	3.462731067	4.13023818	3.9110181
.55	.015790774117	1.60478620245	2.5393714828	3.578617360	4.24696042	3.9180018
.60	.064184754846	1.71113087652	2.7677577032	3.948510566	4.60802246	3.8845042
.65	.148471861894	1.90441739172	3.1952786294	4.649203898	5.24251774	3.5819388
.70	.274995897728	2.21481344618	3.9151534227	5.851104997	6.17293041	2.3028687
.75	.454936423167	2.70147862028	5.1258316257	7.926014789	7.28172233	-2.2928262
.80	.708326301120	3.48732868622	7.2898985394	11.769404481	7.56379527	-19.3280410
.85	1.074194516116	4.86470249400	11.6974912741	19.988202597	-.02639844	-95.6178819
.90	1.642374745726	7.7 131156316	23.2907287643	43.184119969	-84.22129439	-625.5660190
.95	2.705543512080	16.54733324249	73.5308308659	170.623199039	-2,153.75053298	-12,272.7590420

$H_i(p_r: 3, 0)$

p_r \ i	0	1	2	3	4	5
.50	.507949098483	1.85086768363	3.6647898021	6.360611293	9.40919165	11.4188047
.55	.372696539114	1.45350868143	2.8133674491	4.813982250	7.08644580	8.7460757
.60	.266970834511	1.13688226508	2.1622769627	3.661604621	5.35158877	6.6625226
.65	.185041594474	.88060091568	1.6553514866	2.783570025	4.02008609	5.0268526
.70	.122544063162	.67044797554	1.2555332696	2.101578526	2.96766487	3.7485092
.75	.076064469376	.49629567184	.9378311607	1.561263659	2.10448901	2.8067164
.80	.042857509545	.35087324228	.6854212009	1.120422367	1.36024241	2.3292147
.85	.020639458270	.22909542173	.4871465542	.736575726	.68472407	2.9134694
.90	.007414659897	.12792680777	.3362243175	.339103584	.14307829	7.2203932
.95	1.073752507170	4.76001116482	-122.8768387860	-.302174892	1.29590096	32.1537985

$H_i(p_r: 2, 1)$

p_r \ i	0	1	2	3	4	5
.50	zero	-1.59576912160	-3.3080440430	-5.973042695	-9.11603250	-11.2433663
.55	.065078662264	-1.23608334807	-2.5110425703	-4.501628520	-6.86371095	-8.5451594
.60	.105040913358	-.94965140209	-1.9046495071	-3.419356754	-5.19732602	-6.4027912
.65	.125122376755	-.71722563499	-1.430340930	-2.612041842	-3.92587088	-4.6288696
.70	.129377262699	-.52530129234	-1.0714531701	-2.007743789	-2.90817495	-3.0567703
.75	.121086738363	-.3641627780	-.7902252730	-1.559796846	-2.00794013	-1.5384977
.80	.103070560491	-.22677044771	-.5825947009	-1.234400127	-1.02001047	-.0486391
.85	.077984267072	-.10831474081	-.4520808217	-.989300929	.54749381	.3868446
.90	.048730060736	-.00687302560	-.4290512896	-.675320576	4.45812340	-11.9835862
.95	.019386[illegible]17648	.07126406532	-.6400723925	.932492658	21.65952033	-289.0642672

$H_i(p_r: 1, 2)$

p_r \ i	0	1	2	3	4	5
.50	zero	1.25331413732	2.7755917903	5.363197885	8.75961518	11.4214141
.55	.011363755328	.91528794223	1.9981658623	3.944074660	6.66414015	8.9167395
.60	.041328834662	.63626717417	1.3911866797	2.918209506	5.23274521	7.0042971
.65	.084605892039	.39588584704	.9029267955	2.192519242	4.32775274	5.3957768
.70	.136591489394	.17777505947	.5004890450	1.735320938	3.90126255	3.5820853
.75	.192757516453	-.03326656869	.1659328526	1.587061062	3.97663950	.2067634
.80	.247880489382	-.25444627447	-.0993185211	1.928249495	4.58380309	-9.5104411
.85	.294656275912	-.51059925013	-.2471423908	3.344355995	5.17273650	-47.8184443
.90	.320259978490	-.84928606801	-.0423313166	8.138643187	-2.05708485	-276.5227611
.95	.293724162728	-1.40612763997	2.1464886367	32.815363616	-203.47954126	-3989.4200713

$H_i(p_r: 0, 3)$

p_r \ i	0	1	2	3	4	5
.50	zero	zero	zero	zero	zero	zero
.55	.001984289944	.60027649839	1.5637032723	3.2166[illegible]150	5.21091619	6.2170954
.60	.016261021728	1.[illegible]6121036196	3.3302438826	6.924516495	11.21360150	12.7369584
.65	.057209247085	2.[illegible]913483772	5.5684485118	11.801096010	19.07752173	20.1379707
.70	.144207989761	3.10858930559	8.7235719997	19.039304814	30.60637783	27.6997826
.75	.306849954433	4.61173539958	13.6716649866	31.163629217	49.32647713	30.0901260
.80	.596142455455	6.97960725387	22.4470889152	54.569748738	82.856340993	-3.8337716
.85	1.113331242232	11.20910331239	40.8796870830	109.770354544	147.36925693	-287.5572299
.90	2.104788138421	20.46014378819	42.2934716664	293.429537830	228.28634328	-3194.6953042
.95	4.450223106410	51.84470528207	364.8657468717	1661.129658728	-3381.10525075	-102326.1274237

$H_i(p_r: 4, 0)$

p_r \ i	0	1	2	3	4	5
.50	.405284734569	2.77960780095	8.3577107877	18.685798913	34.48894220	52.9242147
.55	.268209287954	2.00022[illegible]12494	6.1103358855	13.482749922	24.65834706	38.1208637
.60	.171903647609	1.42110843325	4.4770029195	9.778325163	17.75861566	27.6433376
.65	.105444957482	.98812111977	3.2697977674	7.077426406	12.79818367	20.0808289
.70	.060868093822	.66450834365	2.3661997484	5.069063370	9.16906928	14.5266829
.75	.032228675174	.42471522687	1.6833556118	3.550810704	6.48879480	10.3588747
.80	.014998088342	.2506[illegible]3	1.1631592324	2.387852994	4.51560362	7.0714193
.85	.005661494095	.129[illegible]	.7630552592	1.491673587	3.09779648	4.0427135
.90	.001445844698	.05168297650	.4498862756	.815127343	2.10454579	-.0434011
.95	.000138912269	.01067465530	.1957751861	.381873836	.91808864	-7.6169875

$H_i(p_r: 3, 1)$

p_r \ i	0	1	2	3	4	5
.50	zero	-1.90985931710	-7.0985931710	-16.968570326	-32.40450532	-50.9309789
.55	.046833549108	-1.27350658145	-5.1115372688	-12.139779912	-22.98280912	-36.4676763
.60	.067636287639	-.81308491594	-3.6891030757	-8.725624823	-16.40077506	-26.3314692
.65	.071300313502	-.48081757394	-2.6571004542	-6.251132396	-11.69023121	-11.1327624
.70	.064262169551	-.24469803752	-1.9021316715	-4.416811870	-8.26658879	-14.0089126
.75	.051304704951	-.08292942482	-1.3471988091	-3.025503118	-5.78156539	-10.3958541
.80	.036069790059	.01937275223	-.9371301786	-1.943120134	-4.06385705	-7.7894122
.85	.021391327130	.07218199465	-.6284991051	-1.081855607	-3.13931148	-5.0221752
.90	.009502269955	.08185335021	-.3798038253	-.421990359	-3.37382709	4.2105984
.95	.002104663727	.05259859924	-.1347308717	-.273928837	-5.01915048	84.5989386

$H_i(p_r: 2, 2)$

p_r \ i	0	1	2	3	4	5
.50	zero	1.00000000000	5.8584073464	15.336828470	30.35317652	48.7541271
.55	.008177872358	.49549536887	4.1348638894	10.892027488	21.30423218	34.5577874
.60	.026611811147	.14438212020	2.93188[illegible]3987	7.775807486	14.97113612	24.7002759
.65	.048212212579	-.09428276923	2.0899372552	5.527684310	10.39306078	17.8873177
.70	.067845502892	-.24694136678	1.5057516327	3.848019277	6.98867134	13.5054309
.75	.081671763915	-.33028075378	1.1087167059	2.520580116	4.43320036	11.6188166
.80	.086746372285	-.35374087242	.8448780859	1.359216912	2.69867510	13.3810144
.85	.080825511238	-.32047661989	.6595326483	.155963986	2.55168910	22.1886589
.90	.062450091911	-.22695952053	.4557807297	-1.365484435	8.73360305	41.0440835
.95	.031887819725	-.06254201543	.1169194886	-2.788407020	65.67337097	-258.5657872

$H_i(p_r: 1, 3)$

p_r \ i	0	1	2	3	4	5
.50	zero	zero	-4.4214141000	-14.137166941	-28.84625478	-46.1688768
.55	.001427984801	.38670335887	-3.3039597074	-10.215900437	-20.17779212	-31.5596007
.60	.010470540552	.63459486475	-2.4050485589	-7.575321921	-13.96098182	-20.75270[illegible]
.65	.032600381788	.77376663574	-1.8853734809	-5.767286323	-9.06852318	-12.2070328
.70	.071628647069	.81809451865	-1.6750345958	-4.467327713	-4.53221939	-5.1257219
.75	.130012969128	.76647876130	-1.7479544483	-3.342897403	.87938711	-.2535435
.80	.208621483299	.59831364221	-2.1157423660	-1.809358183	9.48100274	-5.1592824
.85	.305391651858	.25666491782	-2.8198515667	1.808578541	26.96798150	-70.3966227
.90	.410429718121	-.41007872721	-3.8313734737	14.822439881	68.53239603	-672.3770904
.95	.483133259564	-1.93915680815	-3.4024097097	101.060868525	20.63974156	-14578.2150542

$H_i(p_r: 0, 4)$

p_r \ i	0	1	2	3	4	5
.50	zero	zero	7.4022033008	23.953439248	52.76060206	90.1903681
.55	.000249348547	.15046917602	8.0341915056	25.816436863	56.89459145	96.7324839
.60	.004119682754	.63247450650	10.0841924216	31.938907262	70.57388780	118.0744290
.65	.022043893774	1.55029999801	14.0914407436	44.232796803	98.41968656	160.0717690
.70	.075622743767	3.12892631120	21.2827261144	67.253048372	151.63353676	235.0032202
.75	.206967149124	5.83684664591	34.4251669588	112.001836463	257.89568617	364.3767029
.80	.501726148858	10.72416185695	60.4444481877	208.724467697	495.24380347	561.6863875
.85	1.153893858454	20.52870200014	120.3470406693	461.783721923	1,138.94840673	520.2079610
.90	2.697394805398	44.55937761456	303.1432956970	1,406.291468550	3,598.43976716	-6,402.3369200
.95	7.319965695763	137.86614630113	1,382.6637544975	9,642.008876829	22,267.10018458	-413,221.9426095